DOCTOR OF PHILOSOPHY

Understanding the role of TapA in the formation of B. subtilis biofilms a biochemical and genetic analysis

Earl, Christopher

Award date:
2018

Understanding the role of TapA in the formation of *B. subtilis* biofilms: a biochemical and genetic analysis

Christopher Earl

Thesis submitted for the degree of

Doctor of philosophy

School of Life Sciences

University of Dundee

July 2018

“whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.”

Charles Darwin “On the Origin of Species”, 1859

Declaration

I declare that I am the author of this thesis; that, unless otherwise stated, all references cited have been consulted; that the work of which the thesis is a record has been done by myself, and that it has not been previously accepted for a higher degree. If the thesis is based upon joint research, the nature and extent of my individual contribution is clearly defined.

Chris Earl (candidate)

Prof. Nicola Stanley-Wall (supervisor)

Abbreviations

amp	ampicillin
APS	ammonium persulphate
ATP	adenosine-5'-triphosphate
BMe	β-mercaptoethanol
cDNA	complementary deoxyribonucleic acid
ddH_2O	double distilled water
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
ECL	enhanced chemiluminescence
ECM	extracellular matrix
EGTA	ethylene glycol tetraacetic acid
EPS	exopolysaccharide
et al.	*et alia* (meaning: and others)
GFP	green fluorescent protein
kDa	kilo Dalton
LB	lysogeny broth or Luria-Bertani media
MLS	Macrolides, lincosamides, streptogramines
MSgg	minimal salts glycerol glutamate
mfu	minimal functional unit
MOPs	*(*3-*(N*-morpholino*)*propanesulfonic acid*)*
MW	molecular weight
NMR	nuclear magnetic resonance
OD_{600}	optical density measured at wavelength 600 nm
ORF	open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PHA	polyhydroxyalkanoate
IPTG	isopropyl β-D-1-thiogalactopyranoside
erm	erythromycin
gDNA	genomic DNA
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
RT-PCR	reverse transcription polymerase chain reaction
SDS	sodium dodecyl sulphate
spc	spectinomycin
TAE	tris acetate EDTA
TapA	**T**asA **a**nchoring/assembly **p**rotein; formerly YqxM
TBS	tris buffered saline
TBST	tris buffered saline tween
TCS	two-component system
TEMED	tetramethylethylenediamine
UV	ultra violet
v/v	volume per volume
WT	wild-type
w/v	weight per volume
YFP	yellow fluorescent protein
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Acknowledgements

I would like to thank, first and foremost, the Nicola Stanley-Wall (NSW) lab members for making this thesis possible. When I started in the lab I relied on the help of everyone to learn the experimental techniques required to carry out this project. So thank you to Billy, Sofia, Tanya, Margarita, Yusra, Manisha and Elliot (in no specific order). I would also like to thank the lab for all of their suggestions and discussions which helped to inform and shape the project as it unfolded. Most importantly I would like to thank Prof. Nicola Stanley-Wall for being a great supervisor. Thank you for being very generous with your time and imparting some of your knowledge that I can take forward with me to future endeavours. Nicola along with all members of the NSW group have helped me to re-discover my enthusiasm for science and I have very much enjoyed exploring biofilm formation in *B. subtilis*. I would also like to thank the Professor Cait E. MacPhee group at the University of Edinburgh including Ryan Morris for help with peptide work.

I have been incredibly lucky to work in what must be one of the best microbiology environments in the world in the Division of Molecular Microbiology at Dundee University with great facilities, expertise and services all within walking distance of the lab. A special thank you has to go to Dr. Sarah Murdoch for her excellent training in cloning techniques and experimental design. Also I must thank John Allan, Connor Bowen, Dr. Jon Cherry and Alex Finney for their friendship and scientific discussions which helped my work and I hope helped their work too. My first experience in the lab environment was with Dr. Grant Buchanan who has continued to offer great insight into any cloning project I have under taken. I would like to thank Prof. Frank Sargent and Prof. Tracy

Palmer for going above and beyond the call of duty to develop young scientists through the iGEM competition of which I was lucky enough to take part. Furthermore, their desire to help those in need is admirable and there are many people other than myself who will always appreciate their kindness and support. In truth, nearly all members of the division past and present have been great to work alongside and I thank them for this.

I would like to thank my partner Erin Kenny for her emotional support and patience in what has, at times, been a traumatic experience! No one has been more aware of those difficulties than Erin. Her support meant that I was always able to find some level of enjoyment in what I was doing. I am very lucky to share my life with someone that has that level of patience. A special mention has to go to my very good friend Scott Gallacher who not only helps foster my interest in science but has listened and supported me through the more challenging circumstances as well. Thanks go to Karim Hussain, Liam Johnston, Mehmet Gundogdu, Katie Brown, Ian Brown, Kathleen Brown and anyone else who had to endure my long winded complaints about scientific research! Thanks finally to my Mum, Angela, and my Dad, Stuart, for supporting me regardless of what I wanted to do. I wasn't forced to study at University and carry out a PhD, I did it because of a genuine interest in science and my parents gave me the space required to pursue that interest.

Thanks to all at Dundee Road Runners (DRR) AC. I found solace in running and lifelong friends in DRR.

Publications

Morris, R.J., Schor, M., Gillespie, R.M.C., Ferreira, A.S., Baldauf, L., **Earl, C**., Ostrowski, A., Hobley, L., Bromley, K.M., Sukhodub, T.*, et al.* (2017). Natural variations in the biofilm-associated protein BslA from the genus Bacillus. Scientific Reports *7*, 6730

Erskine, E., Morris, R.J., Schor, M., **Earl, C**., Gillespie, R.M.C., Bromley, K., Sukhodub, T., Clark, L., Fyfe, P.K., Serpell, L.C.*, et al.* (2018). Formation of functional, non-amyloidogenic fibres by recombinant Bacillus subtilis TasA. Molecular Microbiology, *In press*..

Summary

Once thought of as unicellular organisms, it is now appreciated that bacteria can exist as multicellular collectives with emergent properties that distinguish them from their free-living planktonic forms. Biofilm formation is an example of such a multicellular behaviour which is widespread in the prokaryotic world. Biofilms can take diverse forms but this work focusses on the surface-attached biofilm formed by the organism *Bacillus subtilis*. This study explores the properties of the extracellular matrix which gives structure to the biofilm and protects the encased cells from a variety of environmental threats.

The *B. subtilis* biofilm matrix is composed of exopolysaccharide (EPS) and the protein components BslA, TasA and TapA. TasA is a fibre forming protein which has been described in the literature as an amyloid-forming fibre, however, recent work contradicts this finding. The current work focusses on TapA (**T**asA **a**nchoring/assembly **p**rotein; formerly YqxM) which is essential for the development of complex, structured biofilms. TapA has previously been described as a cell-wall associated protein needed for the formation of TasA fibres. TapA secretion has since been demonstrated to involve the action of the signal peptidase SipW with the *tapA-sipW-tasA* genes found together on the *B. subtilis* genome.

Recent work from Erskine *et al* has demonstrated that TasA fibres (fTasA) form spontaneously *in vitro* (Erskine et al., 2018). The findings of this thesis determine that TapA is not needed for the ability of TasA fibres to restore structure when added *ex vivo* to a Δ*tasA sinR* mutant biofilm. Biochemical and genetic analysis of *B. subtilis* biofilms has led to the discovery that TapA is processed to a low molecular weight form *in vivo* and that this process is

dependent on the proteolytic activity of secreted proteases. Subsequently, it was found that a limited part of the N-terminus of the TapA protein is sufficient to restore rugosity to a Δ*tapA* mutant biofilm. Genetic analysis of the minimal functional unit identified key amino acids needed for the function of TapA, highlighting the importance of a potential β-strand secondary structure in the N-terminus of TapA. Evidence is provided that SipW is not required for the ability of exoproteases to process TapA, suggestive that TapA secretion is not SipW-dependent. Experimental evidence indicates that TapA could act as a chaperone facilitating the stability of TasA or as a peptide aiding the formation of TasA fibres *in vivo*.

Table of contents

Table of Figures

List of Tables

1. Introduction

1.1. A brief history of multicellularity in prokaryotes

In a *Scientific American* article of 1988 James A. Shapiro challenged the *status quo* in microbiology by suggesting that bacteria could be thought of as multicellular organisms (Shapiro, 1988). Until this point, bacteria had been described as unicellular autonomous organisms. It has been suggested that this paradigm was a consequence of the adoption of Koch's postulates into medical microbiology more than a century before (Aguilar et al., 2007). Shapiro offered a new and important perspective drawing upon recent research detailing *Chondromyces crocatus* fruiting bodies, cyanobacterial filaments, the swarming behaviour of *Proteus mirablis* and the production of extracellular material by *Pseudomonas putida.* However, the appreciation that bacteria form multicellular structures was not in itself new. In 1877, the same year as Robert Koch published on *Bacillus anthracis* as the causative agent of anthrax, Ferdinand Cohn published hand drawings of multicellular aggregates of *Bacillus subtilis* (Cohn, 1877). It took a further 120 years before the study of *B. subtilis* aggregates, now known as biofilms, began to flourish. The term biofilm was not coined until 1978, by Bill Costerton (Chandki et al., 2011), but the first description of biofilms is attributed to Anton Von Leeuwenhoek in the 1600s. Leeuwenhoek is recognised as the first microbiologist and he used powerful microscopes, which he self-produced, to discover protists and bacteria which he called "animalcules". In a letter in 1684, Leeuwenhoek described microscopic observations of "animals in the scurf of teeth" (Lane, 2015). Today dental plaque is one of the best studied examples of a biofilm.

In the proceeding section I am going to summarise some of the major advancements in our understanding of bacterial multicellularity, including biofilms with a focus on *B. subtilis*. I will then review in detail the existing

understanding we have of the extracellular matrix which provides the structural complexity of the *B. subtilis* biofilm. Finally, I will outline the current model for how *B. subtilis* forms fibres in the biofilm with a focus on the role of the matrix protein TapA in this process.

1.2. Multicellular behaviours exhibited by prokaryotes

Bacteria can and do live as unicellular autonomous organisms but many species spend at least some time as part of complex multicellular communities and some have relinquished unicellular life completely. It is now appreciated that many of these communities contain multiple interacting microbial species (see section **1.3.3**) (Claessen et al., 2014). Multicellularity was previously thought only to be the reserve of eukaryotes such as plants, metazoans, fungi and protists. However, a number of phenomena have been observed in prokaryotes which are indicative of a multicellular lifestyle. These include co-ordinated motility behaviours, cell-cell adhesion, intercellular signalling, cell specialization, matrix production and the formation of complex structures such as biofilms and fruiting bodies. Examples of multicellularity include cyanobacteria such as *Anabaena* species which form filaments where cell-cell adhesion makes them look like “beads on a string” (Kumar et al., 2010) (**Figure 1**A). A fascinating case are the multicellular magnetotactic prokaryotes (MMPs) (**Figure 1**B). MMPs swim as spherical aggregates of between 10 and 40 cells which inhabit aqueous environments. The MMPs have magnetic “organelles” called magnetosomes that are thought to help the MMPs to navigate through magnetic field lines in the water column, possibly in the search for nutrients (Keim et al., 2004b). Interestingly, the division of the entire structure is a concerted behaviour. The other manifestation of multicellular bacteria are aggregates and swarms. Swarms include bacteria such as *Myxococcus xanthus*

with its characteristic mode of motility and fruiting body formation (Claessen et al., 2014). Finally, aggregates include examples of some of the earliest forms of life on Earth.

1.2.1. The origins of multicellularity

Multicellularity has arisen independently in each kingdom of life and the earliest fossil evidence for life is found in the form of multicellular communities of prokaryotes (Grosberg and Strathmann, 2007; Kaiser, 2001). A 2016 publication from Nutman *et al* provided evidence for life forms in 3.7 billion year old structures interpreted to be stromatolites (Nutman et al., 2016). Stromatolites are multi-layered sheets known as microbial mats (normally found in aqueous environments). These mats are thought to be made up of aggregates of photosynthetic organisms, as they are in contemporaneous formations found in Shark Bay, Australia. The concept of aggregates describes cells held together by cell-to-cell contacts both in surface attached communities known as biofilms (**Figure 1**C) and mobile biofilms, also known as flocs, where there is cell-to-cell contact but no surface attachment (**Figure 1**D) (Flemming et al., 2016). It is thought that for microbes, in a natural setting, the biofilm lifestyle is the rule rather than the exception (Hobley et al., 2015). Although much of the focus has been on bacterial biofilms both archaeal and eukaryotic biofilms have been observed (Aguilera et al., 2007; Fröls, 2013).

There is often a distinction made in the literature between complex/true multicellular organisms such as the metazoa and plants and the type of multicellularity exhibited by prokaryotes. It is speculated that prokaryotic cells are by their nature energetically limited and would not, in their present form,

lead to the evolution of “complex life” such as is seen in the animal and plant worlds (Lane and Martin, 2010).

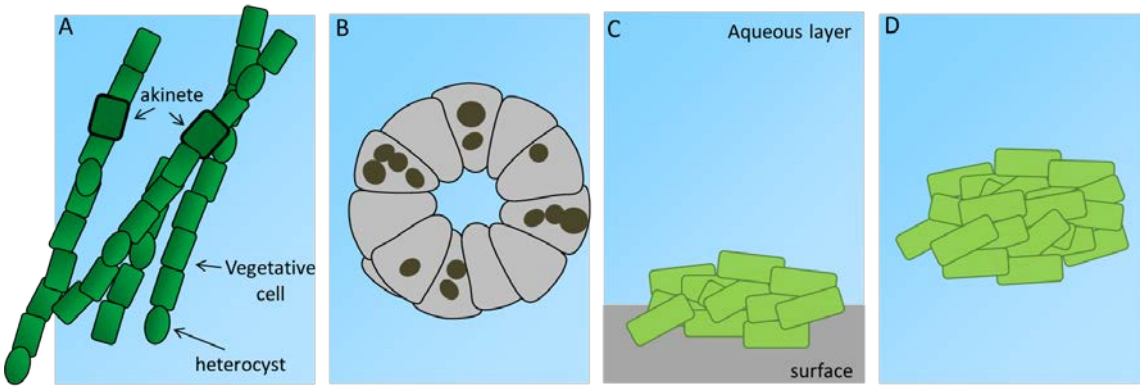


Figure 1: General classes of multicellular bacteria. (A) Filaments formed by the vegetative cells of the cyanobacterial species *Anabaena cylindrical* contain differentiated heterocyst cells specialised for Nitrogen fixation. Cyanobacteria also form cells called akinetes which are thick-walled, dormant cells. **(B)** Multicellular magnetotactic prokaryotes (MMPs) exist as conglomerates with an acellular central compartment around which the cells are arranged. Individual cells contain magnetic iron sulphur crystals. Subcellular structures depicted as dark spheres are thought to be lipid or polyhydroxyalkanoate inclusions (Keim et al., 2004a). In terms of bacterial aggregation, this may be in the form of **(C)** surface-attached communities or **(D)** floating ‘flocs’, both are examples of biofilms found in natural settings often in aqueous environments as depicted here.

1.2.2. The benefits of multicellular lifestyles

There are many benefits that multicellularity conveys that are not available to unicellular organisms. This includes increased size, which may help guard against grazing by predatory microorganisms. As an example, it has been shown using a microbial evolution experiment that the eukaryotic algae *Chlamydomonas reinhardtii* developed multicellular structures in response to the predatory ciliate *Paramecium tetraurelia* (Herron et al., 2018). Other benefits include the co-ordination of cell behaviour by intercellular signalling and cell differentiation where specialisation of cells for particular functions can result in a division of labour and the motility of cells *en masse*.

1.2.3. Cell-cell adhesion

In multicellular communities, cells can be strongly adhered to each other such as in cyanobacterial filaments, or by fimbrae known as the “attachment pilus” found in Gram-negative and Gram-positive bacteria. Otherwise cells can interact by cell-cell interactions mediated by specific receptors and cognate ligands or non-specific molecular forces such as hydrogen bonding and van der waals forces (Dufrêne, 2015). A special example of bacterial adhesion is by the common association of a shared “glue” such as the case in the extracellular polymeric matrix produced by biofilm forming bacteria. Some bacteria such as *Streptomyces* species even resemble fungi in that they produce filamentous, syncytial cells where the branched mycelia contain compartments, which may have multiple chromosomes, divided along the filament by cross walls (Claessen et al., 2014).

1.2.4. Bacterial communication systems

Bacteria can signal between cells of the same species (intraspecies) and with cells of other species (interspecies) signalling. Quorum sensing systems have been demonstrated within and between species of bacteria. Fuqua *et al* coined the term quorum sensing in a landmark review paper of 1994 (Fuqua et al., 1994). The name is derived due to the fact that behaviours regulated by quorum systems are only possible with a bacterial population of a sufficient size. Cell-density-dependent processes are regulated by auto-inducers which in Gram-negative bacteria tend to involve acyl homoserine lactones (AHL). Processes regulated by these systems include the production of light by *Vibrio fischeri* and *Vibrio harveyi* , where quorum sensing was first observed nearly 40 years ago (Nealson and Hastings, 1979) and virulence factor regulation in *Pseudomonas aeruginosa* amongst many others. Since this time many other types of quorum

sensing molecules have been discovered. In Gram-positive bacteria quorum sensing can be regulated by auto-inducing peptide (AIP) systems such as the regulation of genetic competence in *B. subtilis* and *Streptococcus pneumoniae* (possibly the earliest hint of quorum sensing on record) (Tomasz, 1965). To date quorum sensing systems have been implicated in the control of diverse processes such as antibiotic synthesis, exo-enzyme production and the regulation of virulence in human and plant pathogens (Whiteley et al., 2017).

1.2.5. Co-ordinated movement of bacteria *en masse*

Bacterial motility can be thought of as conveying the ability of single cells (or multicellular communities) to either avoid dangers such as toxins, predation, the immune system or to aid in survival and/or the acquisition of nutrients. In the case of MMPs, each cell in the structure has its own flagellum which may be co-ordinated to facilitate movement of the whole organism (Keim et al., 2004a). However, MMPs are a rather special case and bacterial motility can typically be thought of in terms of 5 common modes of movement: swimming, swarming, sliding, gliding and twitching, defined by Jørgen Henrichsen in 1972 (Henrichsen, 1972; Shrout, 2015). First of all, the concept of a swarm i.e. “to move about in great numbers” is important for understanding motility (Kearns, 2010). The predatory multicellular bacteria *Myxococcus xanthus* moves as swarms of cells in concentric ripples towards prey bacteria and it is now thought that this movement is driven mainly by growth rather than chemotaxis (Kaiser and Warrick, 2011). We can use the general term “swarm” to describe this type of cell movement *en masse*. However, the term “swarming” motility specifically refers to cell movement dependent on the rotation of flagella and in this context *M. xanthus* does not display swarming motility. Flagella-dependent movement when occurring in liquid is called swimming motility and it is termed swarming

motility when it is on a surface. In the case of twitching motility this is a slow, jerky “twitching” movement of cells generated by type IV pili which extend and contract (Mattick, 2002). Twitching motility is displayed predominantly by Gram-negative bacteria such as *Pseudomonas aeruginosa* and *Vibrio cholera* (Mattick, 2002). *Myxococcus xanthus* has type IV pili and twitching is one form of motility that contributes to its “swarming behaviour” with the other being gliding (Sliusarenko et al., 2007). In gliding motility, cells move in the direction of the long axis of the cell without the aid of appendages such as pili, fimbrae or flagellum. However, gliding may involve the action of adhesins which temporarily bind to the surface to propel the cells forward in conjunction with the action of bacterial cytoskeletal components or motors. In *Myxococcus xanthus* gliding is aided by the production of a polysaccharide “slime” that may lubricate the area surrounding the cell (Jarrell and McBride, 2008). Sliding, like gliding, is another example of appendage-independent motility. Sliding is dependent on the surface activity of biomolecules self-produced by bacteria allowing them to move passively from an initial point of inoculation (Shrout, 2015).

1.2.6. Cell differentiation

Cyanobacterial filaments, such as those of *Anabaena cylcindrica*, are made up of photosynthetic vegetative cells and specialised heterocysts (**Figure 1**A) (Claessen et al., 2014). The heterocysts do not carry out oxygenic photosynthesis, as they are differentiated to exclude oxygen by a number of physiological adaptations to protect the oxygen-sensitive process of nitrogen fixation which they carry out (Claessen et al., 2014). Additionally, cyanobacterial filaments are often found to contain akinetes (Claessen et al., 2014). Akinetes are thick-walled cells which are resistant to cold temperatures and dessication and develop in response to nutrient limitation. Another example of cell

differentiation in a prokaryote is the asymmetric cell division of *Caulobacter crescentus* (Hughes et al., 2012). This process gives rise to a non-motile stalked cell, which has a holdfast to adhere to a surface, and a “swarmer cell” with a single flagellum (Kearns, 2010; Skerker and Laub, 2004). Many bacteria, including *Bacillus* species and *Clostridium* species, develop endospores which are triggered by nutrient starvation. They are not true “spores”, as they are not offspring as such, but they are highly resistant to a number of stresses and confer survivability in difficult circumstances (McKenney et al., 2012). Historically, these classic examples of bacterial cell differentiation could be determined because of the gross morphological changes that were easily observed. In the last 30 years with the development of more advanced molecular biology techniques we can now look at more subtle changes in gene expression within specific populations of cells. This has shed light on the variety of different roles specialised cells can play. This will be expored in more detail in the context of *B. subtilis* in section **1.5.1**.

To summarise, there have been many advances in understanding the ways in which prokaryotic cells, specifically bacteria, can exhibit behaviours once thought to be the reserve of complex eukaryotes. It has now been shown that bacteria can exhibit specific cell developmental pathways, motility, cell-cell signalling and adhesion. The next section will address the bacterial biofilm as an example of a multicellular community. Biofilms, widespread in nature, can be thought of as arenas in which examples of all of the above processes play out.

1.3. Bacterial biofilms

A biofilm is a complex community of cells which are held together and encased in a self-made extracellular polymeric matrix which can be comprised of protein, polysaccharide, lipids and extracellular DNA (eDNA) (Cairns et al., 2014). The concept of the biofilm matrix can be traced back in the literature to 1935 (Zobell and Allen, 1935). The authors submerged microscope slides *in situ* from the pier at the Scripps institute in California and they quickly became covered with marine bacteria. They speculated that the bacteria formed a “mucilaginous surface” which could form as an attachment point for the planktonic stage of biofouling eukaryotes which attach to ship hulls and submerged man-made structures. Since then it has been thoroughly demonstrated that cells within a biofilm exhibit distinct properties from planktonic forms (Hall-Stoodley et al., 2004).

1.3.1. The societal importance of understanding biofilms

It has been reported by the World Economic Forum in their publication *Global Risks* that antimicrobial resistant infections are a major threat to global health (Blair et al., 2014). Biofilm formation contributes to the antimicrobial tolerance of infections (Van Acker et al., 2014). They can form on medical devices including prosthetic implants used in orthopaedics such as those used in hip replacements and intravascular/urinary catheters (Francolini and Donelli, 2010). Pioneering work in the 1980s looked into the role that biofilm formation played in the context of cystic fibrosis-associated respiratory infections caused by *Pseudomonas aeruginosa* (Lam et al., 1980). There is now great interest into the role that biofilm formation plays in the pathogenic lifestyle of many pathogens such as *Listeria monocytogenes*, Enterohaemorragic *E. coli*, *Salmonella a*nd many others. For these reasons, the development of specific

therapeutics which can disrupt or prevent biofilm formation is now an active area of research (Koo et al., 2017b).

In terms of industrial applications, biofilms have long been utilized for their ability to help process wastewater. By attaching to nutrients and pollutants biofilm-associated microorganisms in wastewater settle out as sludge which can then be removed, a critical step in generating clean drinking water for human populations (Sehar and Naz, 2016). However, biofilms can also incur economic and health costs by attaching to the surface of pipes transporting clean drinking water risking contamination and damage to the pipes (LeChevallier et al., 1987).

Biofilms can be engineered to modify their properties for use in a range of applications, for example, to increase their efficacy in processing wastewater, to build industrially valuable nanomaterials and in the production of renewable energy (Feng et al., 2013; Nguyen, 2017; Wang and Chen, 2009). In many cases, these properties of interest are conveyed by the extracellular polymeric matrix of the biofilm.

1.3.2. The properties of the biofilm matrix

The energetic costs of producing the matrix is considerable (Saville et al., 2011) so what are the benefits of making these often high-molecular weight products?

The selective advantage of a biofilm lifestyle is that biofilms display a number of ‘emergent properties’ (Flemming et al., 2016), in other words, specific properties emerge when bacteria exist as collectives which individual cells do not possess. Matrix production plays a key role in conveying some of these special features, providing structure and stability to the biofilm. Some of the matrix functions discussed in this section are summarised in **Figure 2**. Other emergent features of the biofilm lifestyle include cell specialisation, cells exisiting in distinct

physiological states, motility and regulated cell death. These concepts will be discussed in more detail in section **1.5.1** in the context of *B. subtilis*.

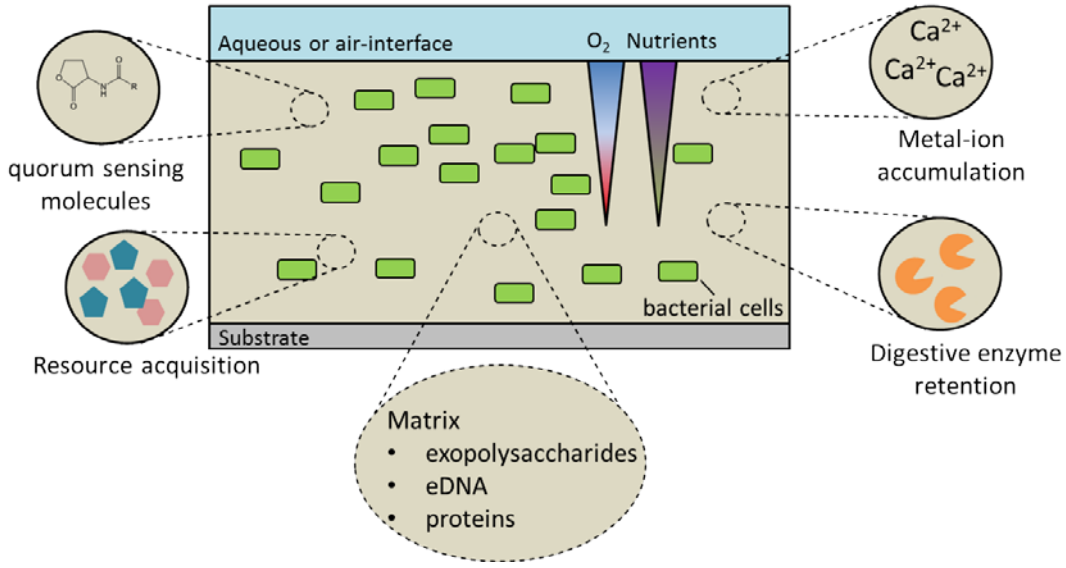


Figure 2: The composition and functions of the biofilm matrix. The biofilm matrix may be composed of exopolysaccharides, proteins, eDNA and lipids. These biopolymers convey structural integrity to the biofilm but can also be important in the acquisition of nutrients and metal ions. Additionally, the matrix can have digestive capacity by the production and retention of a host of degradative enzymes. Quorum sensing molecules have been found to accumulate in the matrix. The matrix displays a number of concentration gradients of cell-cell signalling molecules, nutrients and oxygen which contribute to the heterogeneity of the biofilm environment. The concentration of oxygen and nutrients may become limited in deeper, anaerobic regions of the biofilm towards the substrate.

Matrix components aid in the attachment of cells to surfaces and to each other, forming a 3-dimensional network of polymers that connects cells within the biofilm (Flemming and Wingender, 2010). In this way, the matrix can be thought of as glue that shields the biofilm cells. The different extracellular matrix (ECM) components within the biofilm interact by a number of weak interactions with electrostatic interactions and hydrogen bonding being thought to dominate. This was determined by using the exopolysaccharide of *Pseudomonas aeruginosa* as a model system (Mayer et al., 1999). The interactions between matrix components means that most biofilms display viscoelastic properties (Flemming

and Wingender, 2010). Elasticity is the capacity of a material to return to its original shape and size when stretched or compressed. Viscoelastic materials display elastic properties but also have a viscosity component which can be thought of as “resistance to flow” where a large resistance to flow would be a fluid such as honey by comparison with the low resistance to flow exhibited by water. Under compression *P. aeruginosa* biofilms displayed elastic properties until a break point was achieved after which the biofilm adopts the properties of a viscous fluid (Flemming and Wingender, 2010).

The matrix also helps in resource acquisition by sequestering nutrients and storing water (as much as 97% of the matrix can be water) (Flemming et al., 2016). By retaining water the matrix conveys tolerance to dessication. The physicochemical properties of the matrix also allows for the retention of extracellular enzymes which conveys digestive properties to the matrix (Flemming et al., 2016). Enzymes found in extracellular environments (isolated from marine, freshwater and wastewater sludge) can break down a range of biopolymers including proteins, polysaccharides, nucleic acids and lipids (Flemming and Wingender, 2010). The function of these enzymes may be to; release carbon and energy from larger molecules; aid in cell detachment from the biofilm or as virulence factors important during the infection process of clinically relevant pathogens (Flemming and Wingender, 2010). Interestingly, the biofilm matrix may even protect against the effect of freeze-thaw cycles on marine bacteria isolated from Antarctic environments. A *Pseudoalteromonas* strain was found to secrete a manan-exopolysaccharide which acted as a cryo- and haloprotective substance (Liu et al., 2013).

The sorptive properties of the biofilm matrix can serve to bind and concentrate metal ions. Exopolymeric substances, produced by cyanobacterial, found in stromatolites sequester Ca^{2+} and Mg^{2+} ions from the surrounding seawater (Decho et al., 2005). In some cases calcium has been found to bridge exopolymeric substances to increase biofilm stiffness (Körstgens et al., 2001). For example, it was found that calcium acts by crosslinking the polysaccharide alginate which is one of the main components of the biofilm matrix in the mucoid SG81 strain of *Pseudomonas aeruginosa (Körstgens et al., 2001)*. The sorptive properties of biofilms extend also to toxic compounds. It was found that the exopolymeric substances of biofilms and aggregates used in wastewater treatment can bind to erythromycin and acetaminophen (Métivier et al., 2013).

It has long been known that biofilms demonstrate an increased tolerance to antimicrobials by comparison with planktonic forms of bacteria (Stewart and William Costerton, 2001). The matrix is one of the components which is thought to contribute to these properties. The biofilm matrix may form a diffusion barrier to antibiotic penetration. This is a controversial view which continues to be debated (Van Acker et al., 2014). It has been demonstrated that aminoglycoside antibiotics bind to alginate in alginate-producing strains of *P. aeruginosa* and this impedes their penetration through the matrix (Gordon et al., 1988). In contrast, β-lactam antibiotics did not bind to alginate and penetrated faster (Gordon et al., 1988; Gordon et al., 1991; Nichols et al., 1989). Other polysaccharides of *P. aeruginosa* are thought to interact with antibiotics (Van Acker et al., 2014). The matrix also provides a barrier to the diffusion of disinfectants (Bridier et al., 2011) and to toxic metals. For example, matrix polyssacharides found in *Erwinia amylovora* biofilms complexed with copper to protect against copper stress (Ordax et al., 2010). Due to some of the findings

discussed above the targeted breakdown of the biofilm matrix is now being explored as an approach to aid in the treatment of biofilm based infections and the cleaning of surfaces in hospital environments to remove potential pathogens.

Biofilms can also facilitate intraspecies and interspecies signalling between cells by increasing the localized concentration of chemical signals. It has been found that the concentration of quorum sensing molecules such as acyl homoserine lactones (AHL) can be up to 1,000 fold higher within biofilms by comparison with environments occupied by free-living cells (Charlton et al., 2000; Hense et al., 2007). It has been shown that the Fap fibrils of *P. aeruginosa* biofilms bind to quorum sensing molecules and this is thought to play a role in determining the localised concentration of signalling molecules within the matrix (Seviour et al., 2015).

High cell density, the presence of extracellular DNA (eDNA), and the increased level of competence may all contribute to the exchange of genetic material within biofilms (Madsen et al., 2012). The role of eDNA has been established in the biofilms of a number of species. By treating biofilms at different stages with DNase to degrade extracellular DNA it has been shown that eDNA seems to be important for the structure of immature biofilms. The biofilms of *S. epidermidis* (Qin et al., 2007), *V. cholerae* (Seper et al., 2011) and *P. aeruginosa* (Whitchurch et al., 2002) can be disrupted by DNase treatment but the effectiveness of this declines as the biofilm ages. It is thought that the structural role of eDNA is adopted by other matrix components later in development (Okshevsky et al., 2015). The function of eDNA is thought to be mainly structural, acting as an attachment point for cells and other matrix components,

additionally, there is now interest in the conductive role of eDNA in transferring electrons through the biofilm (Das et al., 2013). Finally, in some cases it has been shown that DNA can serve to prevent aggregation and attachment. In the case of the marine microorganism *Phaeobacter inhibens* it is thought that this process allows cells to leave a hostile habitat which is causing the lysis of siblings (Segev et al., 2015).

1.3.3. Biofilms in nature exist as multispecies communities

Much of what is known regarding the biofilm matrix is a result of studies using single-species model systems to study specific matrix components. Historically, the characterisation of microbial communities was dependent upon culture, microscopic and PCR based methods. However, advancements in culture-independent approaches has led to the discovery and detection of diverse microbial species that previously went unnoticed. Approaches such as high-throughput next generation sequencing platforms like Illumina allow for a more detailed description of microbial community composition and has led to the ever greater appreciation of the polymicrobial nature of biofilm communities and microbial communities generally (Wolcott et al., 2013). Dental plaque is a classical example of a multi-species biofilm consisting of hundreds of different species including bacteria and yeasts like the biofilm-forming *Candida albicans* (Ereshefsky and Pedroso, 2015).

As the complexity of our understanding of natural biofilms increases much remains to be understood about the diverse processes at play within these living structures. For this purpose, monospecies biofilms produced by organisms such as *B. subtilis* are a mainstay in research. The study of *B.*

subtilis has led to a clearer understanding of the broad principles governing biofilm formation and informs the development of biofilm-specific therapeutics.

1.4. General introduction to *B. subtilis*

Bacillus is a genus of rod-shaped, Gram-positive bacteria found within the phylum Firmicutes. *Bacillus* species include pathogens of humans and animals such as *Bacillus anthracis,* the causative agent of anthrax (Moayeri et al., 2015), and *Bacillus cereus* which causes food poisoning in humans often after the consumption of rice dishes (Bottone, 2010). Once cooked, rice which has been kept at room temperature for prolonged periods of time carries an increased risk of causing foodbourne illness. This is due to the formation of endospores by *B. cereus* which are resistant to killing during the boiling of rice and which can then germinate in prepared rice. The formation of resistant, dormant endospores in response to nutritional and environmental stress is a common feature of *Bacillus* species (Earl et al., 2008). *B. subtilis* is non-pathogenic and for this reason has served as a model organism for studying *Bacillus* species.

B. subtilis is a mesophilic organism (with optimal growth temperatures in the moderate range of 30-37^{o}C). This likely places *B. subtilis* in the very top layer of soil, the O layer, which is the organic matter layer and the A layer, just underneath, known as the surface soil with the most accumulated organic matter (Hillel and Hatfield, 2005). *B. subtilis* produces a positive result in the catalase test indicating that it can inhabit an oxygenated environment (O'Donnell et al., 1980), the catalase enzyme is required to convert the toxic oxygen metabolite, hydrogen peroxide (H_2O_2), to oxygen and water. When key nutrients such as carbon, phosphorus or nitrogen become limited then *B.*

subtilis will produce resistant endospores (Chubukov and Sauer, 2014). *B. subtilis* has been isolated from a range of plants and is found in greater abundance than other spore forming bacteria in the rhizosphere of plants (Earl et al., 2008). The species *B. subtilis* is found ubiquitously in the soil environment across the globe from the Mojave dessert to more temperate climes (Roberts and Cohan, 1995). Due to its genetic tractability and ease of cultivation *B. subtilis* has become a model organism for understanding *Bacillus* species and the Firmicutes more broadly. *B. subtilis* has become the archetypal model organism to elucidate fundamental properties of Gram-positive bacteria. For this reason a host of genetic and computational tools (Mäder et al., 2011; Michna et al., 2013) have been available for a significant length of time. The *B. subtilis* 168 strain was one of the earliest bacterial genomes to be sequenced (Kunst et al., 1997).

1.4.1. *B. subtilis* use in biotechnology

B. subtilis species are known to produce a host of antifungal and antibacterial compounds with an estimated 4-5% of a typical *B. subtilis* genome being devoted to antibiotic production (Stein, 2005). For example, *B. subtilis* produces a surfactant known as surfactin which is a cyclic lipopeptide with antimicrobial properties (Schor et al., 2016). Across the many hundreds of *B. subtilis* isolates obtained, there are upwards of 24 different antibiotics discovered (Stein, 2005). Antibiotics produced by *Bacillus* species that have been utilised in medical treatments include gramycidin S, bacitracin, polymyxin and tyrotricidin (Yilmaz et al., 2006).

B. subtilis has been shown to promote plant growth and this can be dependent on biofilm formation (Bais et al., 2004). There are now a number of studies

which have investigated *B. subtilis* biofilm formation in the context of plant roots (Beauregard et al., 2013; Chen et al., 2012; Chen et al., 2013). The growth promoting effects may also be due to the elicitation by the bacterium of induced systemic resistance in plants (Kloepper et al., 2004) and production of toxic compounds which may exclude pathogen growth (Nagórska et al., 2007). Naturally occurring *Bacillus* strains have been utilized for their bactericidal and fungicidal properties to prevent diseases of food crops such as blight and scab. One approach is to apply bacteria-containing formulations as seed treatments or sprays directly onto plants. It is thought that the biocontrol properties could arise from competition with pathogens for nutrients or attachment sites within the plant niche or by the production of antibacterial and antifungal compounds.

As *B. subtilis* is naturally competent, genetically amenable and a prolific secretor of enzymes, its use in biotechnology applications is widespread and of great economic importance (van Dijl and Hecker, 2013). *Bacillus* species have great fermentation properties and can produce high yields of proteins and vitamins (Nicolas et al., 2012). They have the additional benefit of not producing significant toxic products. Applications of alkaline proteases originating from *B. subtilis* include their use as: components of detergent formulations and in the production of biomass from natural waste (Anwar and Saleemuddin, 1998). These enzymes have broad protease activity often replacing expensive and environmentally harmful chemicals in these processes. Subtilisin encoded by the gene *aprE* and other *Bacillus* proteases have been included in commercial laundry detergents and the industry continues to explore protein variants selecting for the ideal properties for specific applications (Anwar and Saleemuddin, 1998). *B. subtilis* also produces lipases and amylases which in addition to proteases aid in the breakdown of complex biological material in the

process of wastewater treatment. A number of *B. subtilis* products are approved for use as additives in food products already available. *B. subtilis* grown on cane molasses has been shown to produce polyhydroxyalkanoates (PHAs). The PHAs serve as a carbon/energy store for the bacteria that produce them. Interestingly, PHAs can be converted into bioplastics. Due to their biodegradability these are an exciting, renewable alternative to existing plastics (Lu et al., 2009). The *B. subtilis var. natto* is used to ferment soybeans into the "stringy" Japanese delicacy of natto. *B. subtilis* alongside other *Bacillus species* have been detected in the faeces and ileal biopsies of human volunteers (Fakhry et al., 2008) and *B. subtilis* is also approved as a dietary probiotic (Elshaghabee et al., 2017). It is taken in the spore form which is thought to help survival through the harsh environment of the gastro-intestinal tract such as in the highly acidic stomach environment. Due to the survivability of spores their utility as a vehicle for vaccine antigen presentation in the host has been explored (Duc et al., 2003). Like most probiotics there remains controversy as to the actual health benefits of consumption but a mechanism has been proposed where *B. subtilis* can antagonise the growth of pathogens through the production of any of a number of antimicrobial products (Elshaghabee et al., 2017).

1.5. Studying the *B. subtilis* biofilm

B. subtilis biofilms in the laboratory are typically grown such that they are composed of a genetically identical (isogenic) population of bacterial cells. On agar, the biofilms form at the air-surface interface (**Figure 3**A), and in liquid, the bacteria form a pellicle biofilm at the liquid-air interface (**Figure 3**B). There are a number of different media which are known to promote biofilm formation such as E medium, 2x SGG (a rich medium) and MSgg (Minimal Salts glycerol glutamate), although the mechanism by which biofilm formation is achieved is not well understood (Shemesh and Chai, 2013). Recently, it was found that the presence of glycerol and manganese could facilitate the development of biofilms and pellicles in a modified LB (lysogeny broth) (Shemesh and Chai, 2013). This indicates that the presence of glycerol and manganese as components in MSgg medium may be a key factor in the ability of this medium to promote biofilm formation.

It was shown that “domesticated” laboratory strains of *B. subtilis*: 168, PY79 and JH642, do not typically form structured pellicles in MSgg media (Branda et al., 2001). However, a more recent study compared 168 stocks originating from multiple laboratories and found that while some were unable to form complex colonies others developed robust biofilm structures (Gallegos-Monterrosa et al., 2016). Additionally, it was found that the soil bacterium *Lysinibacillus fusiformis* M5 can induce wrinkle formation, in strain 168, via the production of a diffusible signalling molecule (hyopoxanthine) (Gallegos-Monterrosa et al., 2017). For the most part, the study of biofilm formation utilizes “undomesticated” natural isolates which form highly structured pellicles with a “vein-like appearance”. When spotted onto semi-solid MSgg agar plates, the natural isolates form colonies with a rugose architecture (Branda et al., 2001). This includes the *B.*

subtilis strain denoted NCIB3610 (Genome accession number: NZ_CM000488), previously known as “Marburg”. Details of the isolation of this strain remain mysterious and were first reported in 1930 (Conn, 1930). Strain 168 is a genetically competent derivative of parental strain NCIB3610 generated from x-ray or ultra-violet (UV) light mutagenesis (Zeigler et al., 2008). Strain PY79 is related to 168 with both strains sharing a more recent common ancestor than the parental NCIB3610 strain.

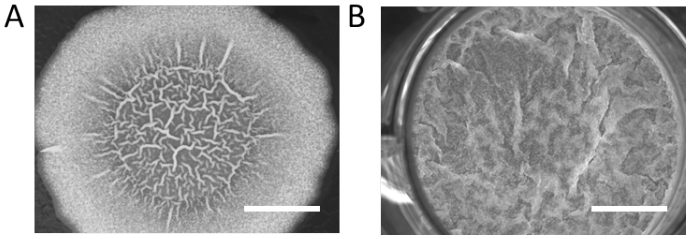


Figure 3: Biofilm formation as displayed by the wild-type undomesticated *B. subtilis* strain NCIB3610. (A) This image demonstrates the “wrinkled” morphology of colonies grown on MSgg agar plates, at an surface-air interface. **(B)** An image of a pellicle, NCIB3610 displays a rugose structure as it grows in MSgg broth, in this case at an air-liquid interface. Thanks to Elliot Erskine for kindly providing this pellicle image. Scale bar = 0.5 cm.

1.5.1. Biofilms as complex microbial communities

Despite the isogenic nature of *B. subtilis* biofilms, studied in laboratory conditions, there are a striking range of gene expression profiles on display within the structure (Lopez et al., 2009). This is testament to the behavioural plasticity of *B. subtilis* and alludes to the diversity of microenvironments that defined cell populations can find themselves in. This is generated by differences in nutrient availability, oxygen gradients, signal concentrations, proximity to the surface and other cells. These variables are dynamic, changing through space and time (spatio-temporal) with biofilm formation seemingly following a pre-determined genetically encoded program. Although whether this qualifies as “microbial development” in a process analogous to development in higher

organisms remains contentious (Monds and O'Toole, 2009). The variety of behaviours exhibited within the *B. subtilis* biofilm are summarised in **Figure 4**, and I will discuss these in more detail in the following section.

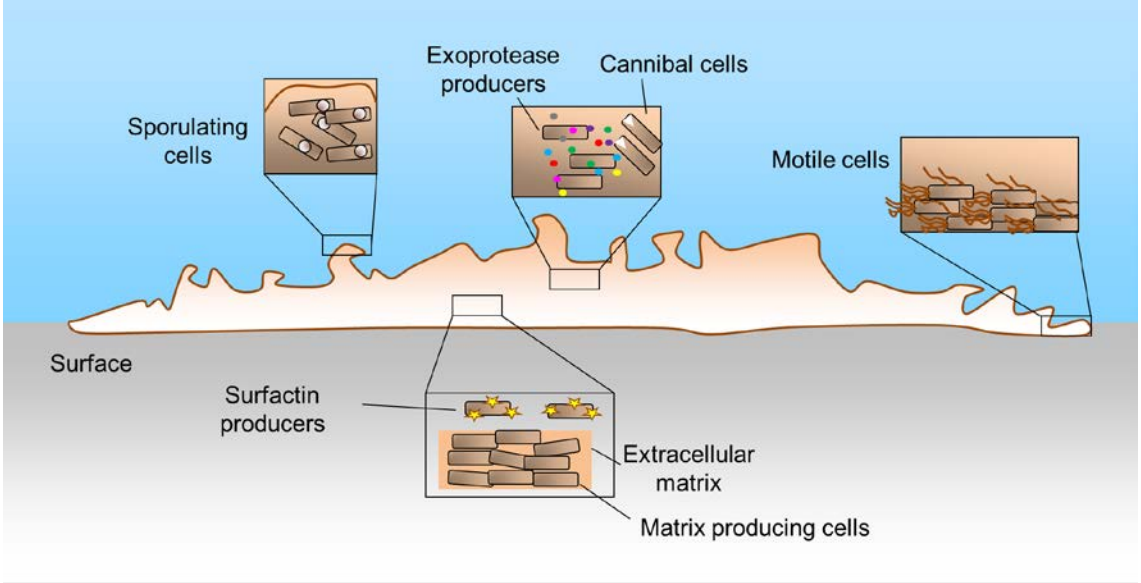


Figure 4: *Bacillus subtilils* biofilms display cells with multiple different gene expression profiles leading to specialization. Sporulating cells are localized at the top of aerial projections. It has been demonstrated that matrix producing cells and exoprotease producing cells form separate populations which are not mutually exclusive as matrix cells have the potential to become exoprotease producers and *vice versa*. Motility within the biofilm is dependent on motile cells which display surface flagella leading to swarming motility. Additionally, surfactin producing cells allow for the movement of cells by sliding motility. Cannibal cells produce toxins which results in the killing of siblings to release nutrients and delays entry into the energetically expensive process of sporulation. This figure is adapted from (Mielich-Süss and Lopez, 2015).

1.5.2. Sporulating cells

The heterogeneous organisation of cell phenotypes in the biofilm was demonstrated clearly when it was discovered that endospore formation was developmentally regulated within the *B. subtilis* NCIB3610 biofilm (Branda et al., 2001). The authors generated a construct of the spore-associated gene *sspE* fused to *lacZ* and this was integrated into the genome. The resulting strain was then grown on MSgg agar containing X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside). In this way, the authors could locate cells within the biofilm

which were expressing sporulation genes. Sporulating cells turned blue because *lacZ* encodes a β-galactosidase that breaks down X-gal to produce a blue pigment. These cells were found to be localised at the top of aerial projections of NCIB3610 pellicles (**Figure 5**). In this same publication, the authors discovered that mutants lacking the *spo0A* gene, which encodes the key transcriptional regulator governing entry into sporulation, could no longer form biofilms with wild-type structural complexity (Branda et al., 2001).

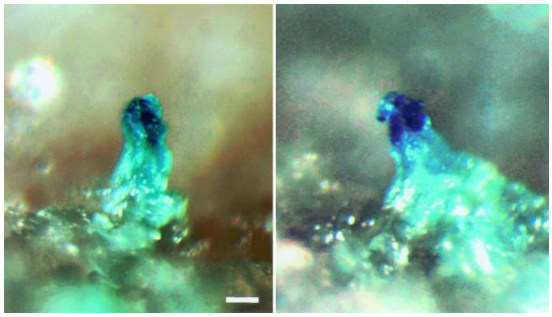


Figure 5: Sporulation genes are expressed in cells localized to the tips of aerial projections of *B. subtilis* NCIB3610 pellicle biofilms. The sporulation gene *sspE* was fused to the *lacZ* gene and this construct was integrated into the genome to create a reporter strain. The *lacZ* gene encodes a β-galactosidase which breaks down X-gal present in the medium to produce a blue compound. In this case, the blue compound is observed under conditions in which the *sspE* gene, and thus *lacZ*, is expressed. This image is reproduced from published work (Branda et al., 2001). Scale bar = 50 μM.

1.5.3. A complex regulatory network controls matrix production

Much light has now been shed on the complex regulatory pathways that control biofilm formation in *B. subtilis*. Since matrix production is an energetically expensive process it is controlled by a tightly regulated genetic program. Specifically, SinR is known as a “master” regulator of biofilm formation functioning as a negative transcriptional regulator of the matrix genes *tapA-sipW-tasA* and the *eps* operon (Chu et al., 2006; Kearns et al., 2005). SinR function is in turn antagonised by the action of SinI (Kearns et al., 2005). In brief, when there is a low-level of Spo0A phosphorylation it can activate the

expression of *sinI*, the SinI protein then sequesters SinR relieving the inhibition of matrix gene expression (**Figure 6**A). Moreover, there is a double-negative feedback loop involving the *slrR* gene, and the SlrR and SinR proteins (Vlamakis et al., 2013). When SlrR levels are low, SinR represses both matrix gene and *slrR* expression. However, when SlrR levels are high it sequesters SinR in a complex which relieves the inhibition of matrix gene expression and of *slrR* expression (Chu et al., 2008) (**Figure 6**A). For these reasons, the low-SlrR and high-SlrR states are self-reinforcing. Finally, another pleiotropic regulator called AbrB was found to negatively regulate the *tapA-sipw-tasA* operon (Stöver and Driks, 1999a), in addition to repressing the *eps* operon (Chu et al., 2008) and *bslA* (Verhamme et al., 2009). An intricate signalling network is found upstream of these interactions adding further complexity to the regulation of biofilm formation. There are a number of environmental signals which can influence the phosphorylation status of Spo0A with 5 distinct sensor kinases inputting phosphate into the phosphorelay system upstream of Spo0A (Aguilar et al., 2010) (**Figure 6**A).

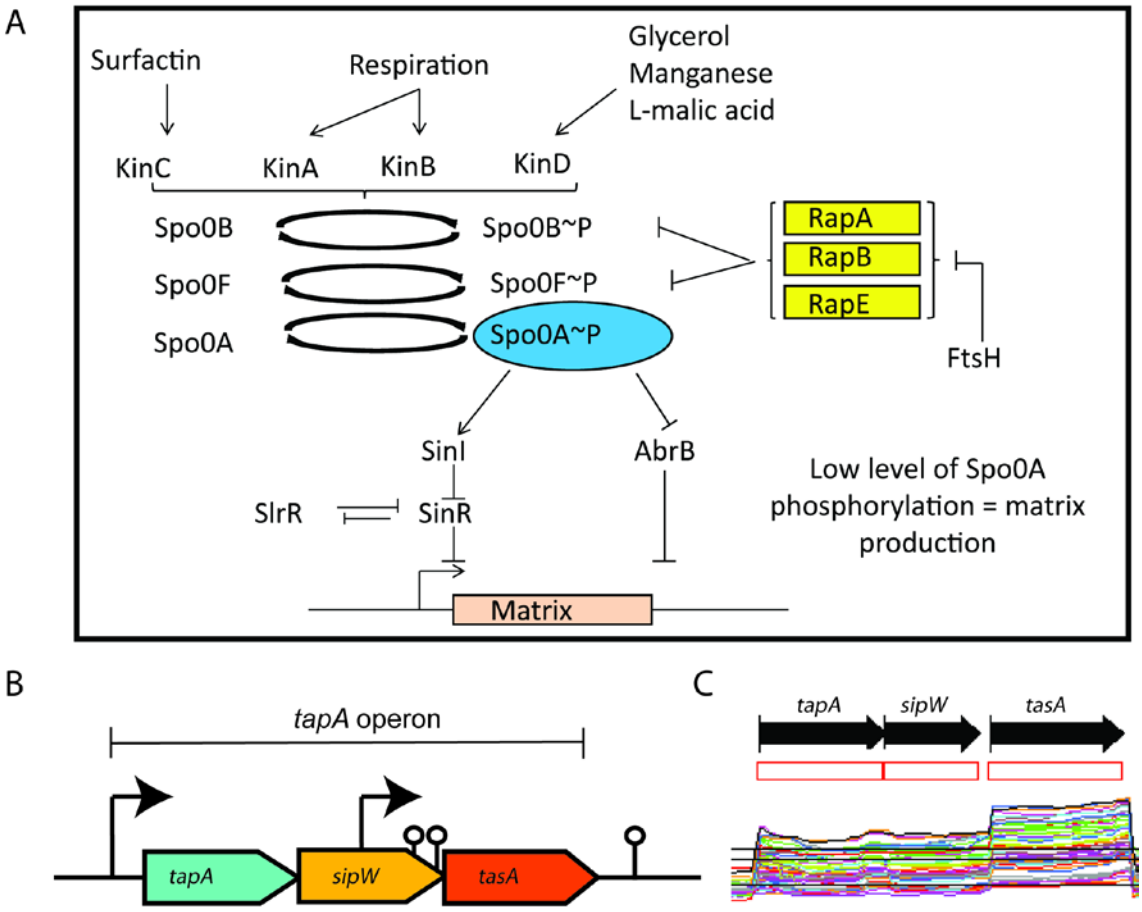


Figure 6: There is a complex regulatory network which controls the expression of the biofilm matrix genes. (A) A number of environmental signals including the presence of surfactin, glycerol, manganese and L-malic acid are detected by the Kin family of two-component systems. Activation of these systems results in the phosphorylation of downstream targets ultimately influencing the level of Spo0A phosphorylation which is a response regulator controlling the expression of genes whose gene products are involved in matrix operon expression levels. A low-level of Spo0A phosphorylation results in matrix production whereas high-levels are linked to sporulation. The Rap family of protein phosphatases regulate the phosphorylation levels of Spo0B and Spo0F response regulators. Additionally, there are many other components feeding into this pathway that are not shown. This figure is adapted from (Mielich-Süss and Lopez, 2015). **(B)** The matrix operon or *tapA* operon is composed of the genes *tapA-sipW-tasA* with *tapA* promoter (arrow) driving expression and a putative promoter internal to the *sipW* gene giving additional control of *tasA* expression. Transcriptional terminators (lollipop) were predicted using the ARNold online server (see supplemental **Figure S6**). **(C)** The presence of an internal promoter within the *sipW* coding region is inferred by the increased level of *tasA* transcript levels, with respect to *tapA* and *sipW*, as shown by global transcriptome analysis of *B. subtilis* (Nicolas et al., 2012). The *tapA* operon is shown in black and the transcriptional units are shown in red, below which the transcription profile of each gene is shown for *B. subtilis* grown in a number of different conditions (represented by differentially coloured lines).

1.5.4. The expression of the *tapA* operon is bistable

The *tapA* gene is present in the *tapA-sipW-tasA* operon (**Figure 6**B and **Figure 6**C). Many studies have used transcriptional fusions utilizing the *tapA* promoter to control the expression of fluorescent proteins such as GFP (green fluorescent protein) as a read-out of matrix gene expression (Chai et al., 2008). Using this method it has been found that matrix gene expression is not found in all cells of the biofilm but in specialized matrix producing cells (Chai et al., 2008). This stable expression of genes in a sub-population of an isogenic community with no detectable expression of the same genes in the other cells of the biofilm is termed bistability (Dubnau and Losick, 2006). An important study used transcriptional reporters consisting of promoters known to control genes involved in matrix production (*PtapA*), motility (*Phag*) or sporulation (*PsspB*) fused to genes encoding the fluorescent protein YFP (yellow fluorescent protein) (Vlamakis et al., 2008). This allowed a visualisation of different gene expression profiles in samples from biofilms which were sectioned and observed using microscopy. It was found that there were well defined zones of expression for each of the promoters. At the 12 h time point most of the cells were motile but at later stages this signal was reduced and mostly restricted to the base and edge of the biofilm (Vlamakis et al., 2008). However, matrix producing cells were found as patches throughout the biofilm. As was shown previously sporulation cells were found in aerial structures and this occurred at a late stage (72 h) of biofilm formation. It was found that motile cells tend to become matrix cells over time within the biofilm and that sporulating cells tend to arise from non-motile community members (Vlamakis et al., 2008). The authors looked at the localisation of the three distinct cell types in cells which do not produce a normal extracellular matrix by using a *tasA* mutant which forms

flat, featureless biofilms. Using this method it was found that motile cells were no longer localised to the base of the biofilm, there was an increase in matrix operon expression and a dramatic decrease in sporulation.

This work has since been expanded, with a *tapA* reporter used to detect the localisation of *tapA* expression in the context of whole complex colonies grown on MSgg agar (Srinivasan et al., 2018). During the first 24 h of biofilm development *tapA* expression is high throughout the colony. As the colony matures *tapA* expression becomes localised to the periphery. The presence of localised *tapA* expression near the leading edge is a result of reduced *tapA* expression in the biofilm interior over time. Cells transition from *tapA* expression to sporulation in the vicinity of this propagating front with increasing levels of *sspB* expression observed at later stages. In late stage biofilms, sporulating cells dominate the biofilm edge but the authors detect a thin layer of matrix producing cells below them. They state in conclusion that nutrient-availability is the main factor determining matrix production (Srinivasan et al., 2018).

Importantly, it has now been shown that in some conditions *tasA* is expressed at a higher abundance than *tapA* and *sipW* and there is now experimental and bioinformatic evidence that the *sipW* coding region contains a promoter that may influence *tasA* expression (Nicolas et al., 2012). In support of this, a *tasA* deletion strain could only be complemented to wild-type when introducing the *tasA* gene with the upstream *sipW* sequence (Erskine et al., 2018). These recent findings should be kept in mind when discussing studies describing the expression of the matrix operon which utilise transcriptional fusions of the *tapA* promoter with a fluorescent read-out.

1.5.5. Motile cells

B. subtilis strains can exhibit swimming, swarming and sliding motility (Hughes and Kearns, 2017). Swimming motility occurs in a liquid environment whereas swarming occurs on a surface (Henrichsen, 1972). Both forms of movement are dependent on flagellar activity. The ability of bacteria to swim can be assessed by growth on 0.4% (w/v) LB agar, for swimming, or in LB solidified with 0.8% (w/v) agar for swarming (Cairns et al., 2015). Swarming motility is a fast movement and in the case of *B. subtilis* NCIB3610 it requires the production of the biosurfactant surfactin (Arima et al., 1968; Kearns and Losick, 2003).

Time-lapse microscopy has been used to understand the role of motility in biofilm formation. Fluorescent reporters were used to visualize matrix expression (using cells expressing *PtapA-yfp*) and flagella expression (*Phag-cfp*). Consequentially, expression from one of the two promoters occurred in any single cell but not together, or in some cases not at all (Vlamakis et al., 2008). The proportion of motile cells in the population decreases from 12h, at 70%, to a level of 9% at 72h. These results demonstrated that matrix production and motility are spatiotemporally regulated within the *B. subtilis* biofilm.

Sliding motility is flagella-independent and has been described as surface translocation generated by expansive forces in a growing colony (Fall et al., 2006). This is thought to be achieved by the lowering of friction between the bacterial cells and the surface (Fall et al., 2006). *B. subtilis* has been shown to display sliding motility which was demonstrated using a flagella-less *hag* mutant (Fall et al., 2006)**.** Sliding motility has been shown to be dependent on two distinct cell types: surfactin producers and matrix producers. These sub-

populations combine to aid the formation of chains of cells which tightly align to form “Van Gogh” bundles (van Gestel et al., 2015). By this mechanism, these chains can then push away from the colony.

To summarise, *B. subtilis* exhibits a range of motility behaviours including flagella-independent sliding and flagella-dependent swarming. Motile *hag* expressing (flagellar) cells are distinct from the population of cells expressing the matrix promoter. Finally, matrix producing cells and surfactin producing cells combine to generate sliding motility.

1.5.6. DegU controls motility and enzyme production

It is thought that *B. subtilis* has developed an arsenal of macromolecule degrading ezymes as a survival mechanism to release simpler molecules when nutrients become scarce, for example, during the stationary phase of growth in liquid culture (Msadek, 1999). The *B. subtilis* genome encodes at least 7 secreted proteases and 2 cell wall associated proteases. The two component signal transduction system DegS-DegU controls expression of some of the exoprotease genes (Dahl et al., 1992; Ogura et al., 2001; Tanaka et al., 1991). High levels of DegU phosphorylation result in exoprotease production but inhibit motility (Amati et al., 2004). The expression of *bacillopeptidase* (*bpr*) gene was observed at a low level in cells in the biofilm at 17 h but was found to increase over time until late on in biofilm formation (72 hours for biofilms and 96 hours for pellicles) (Marlow et al., 2014). A population of highly fluorescent cells were found towards the surface of the biofilm and the exoprotease producing cell population overlapped with matrix producing cells. Also these were not terminally differentiated cell fates with exoprotease cells becoming matrix cells and vice versa (Marlow et al., 2014). In support of this result, the supernatant

collected from biofilms at later timepoints had higher exoprotease activity than that taken from immature biofilms (Marlow et al., 2014).

1.5.7. Cannibalism and cell death in *B. subtilis* biofilms

For *B. subtilis*, sporulation is a response to nutrient deprivation but the process is energy intensive and so not always desirable. In some cases *B. subtilis* can delay sporulation by directing the killing of other cells in the biofilm to release nutrients. So-called “cannibal” cells produce and secrete the two toxins Skf and Sdp, which are 26 amino acid and 42 amino acid peptides respectively. When added to planktonic cultures of *B. subtilis* only Sdp induced killing, although on spot test assays both showed killing effects, with Sdp being more potent (Liu et al., 2010). Additionally, cannibal cells are resistant to the effects of the toxins (González-Pastor, 2011; González-Pastor et al., 2003). As with matrix production, cannibalism is activated by low-levels of Spo0A phosphorylation leading to the expression of the two toxins (Fujita et al., 2005). For this reason, matrix and cannibalism populations overlap with many matrix cells also producing toxins (López et al., 2009). As a result of this matrix cells reduce the population of non-matrix producing cells by releasing toxins.

In this section I have reviewed the heterogeneity of gene expression within the *B. subtilis* biofilm, which is indicative of a potential division of labour. This is exemplified by the sharing of TasA and EPS as demonstrated by the restoration of biofilm complexity upon mixing of *eps* and *tasA* mutant cells, which would otherwise display featureless biofilms (Dragoš et al., 2018). Given this understanding I am now going to discuss the properties of the biofilm matrix.

1.6. The extrapolymeric matrix of *B. subtilis* biofilms

1.6.1. Cells in the matrix are connected by a network of channels

It was found that to transport nutrients from regions of high concentration to regions where nutrients are low the *B. subtilis* matrix contains a network of channels (Wilking et al., 2013). These channels have a low resistance to the flow of liquids and were found to form when *B. subtilis* is grown on agar (Wilking et al., 2013). The wrinkles of the *B. subtilis* biofilm contribute to the formation of these channels (Wilking et al., 2013). Another method of transport that is possible in *B. subtilis* biofilms is by the secretion of extracellular vesicles. It was found that *B. subtilis* 3610 produced vesicles when grown as pellicles on MSgg media although the content of vesicles and their function remains unclear (Brown et al., 2014).

1.6.2. The function and composition of the matrix

The *B. subtilis* matrix consists of an exopolysaccharide (EPS) (Branda et al., 2001) and the protein components BslA, TasA and TapA (Branda et al., 2006; Romero et al., 2011). Deletion of the genes required for the production of the main biofilm matrix components *epsA-O*, *bslA*, *tapA* or *tasA* results in an abolishment of rugose biofilm morphology. TasA is the dominant protein in the *B. subtilis* biofilm (Branda et al., 2006) and TapA (formerly YqxM) has been reported to be required for the assembly of TasA fibres (fTasA) (Romero et al., 2014). I will discuss TapA and TasA in detail in **section 1.7**.

1.6.3. BslA forms the bacterial “raincoat”

An interesting charactersitic of *B. subtilis* biofilms is that the surface of the biofilm is extremely non-wetting, limiting the penetration of water, liquid antimicrobials and even gases (Epstein et al., 2011). Given this finding a

number of studies followed into the mechanisms which convey these properties. It was found that BslA (Biofilm surface layer protein A, formerly YuaB) is an amphiphilic protein which contributes to the surface repellancy by forming a hydrophobic layer on the surface of the biofilm (Hobley et al., 2013; Kobayashi and Iwano, 2012). Also it was shown that BslA is a communal product which is shared within the biofilm (Ostrowski et al., 2011). The presence of a layer of BslA at the periphery of the biofilm also contributes to the physicomechanical properties of the biofilm by impacting on the roughness and surface stiffness (Kesel et al., 2016). This is illustrated by the flat, featureless biofilms formed by Δ*bslA* mutants (Ostrowski et al., 2011). BslA has been described as bi-functional, where the dimer form of the protein is required to convey the hydrophobic properties of the biofilm and the monomeric form is needed for the structural integrity of biofilms (Arnaouteli et al., 2017). The BslA protein demonstrates an intriguing approach to remaining soluble despite its hydrophobic properties. In an aqueous environment the hydrophobic cap remains unstructured, however, when the protein is at an air-water interface the hydrophobic side chains re-orientate towards the air (Bromley et al., 2015). These unique properties of BslA has led to interest in its use a stabilizer of emulsions in the food and cosmetic industries (Stanley-Wall and MacPhee, 2015). It has been hypothesised that the water repellancy of *B. subtilis* biofilms may be an important adaptation to the plant root environment it inhabits. It may function by protecting the bacteria from rainwater or from antimicrobials produced by competing bacteria in the soil (Arnaouteli et al., 2016).

The *eps* operon is a 15 gene operon encoding putative biosynthesis and assembly enzymes for the production of exopolysaccharide. The deletion of the *eps* operon resulted in the formation of biofilms which were flat and entirely

wetting (Branda et al., 2001; Epstein et al., 2011; Kobayashi and Iwano, 2012). The loss of biofilm hydrophobicity in the *eps* mutant is predicted to be due to a loss of an important EPS to BslA interaction (Arnaouteli et al., 2016), as BslA is released into the culture supernatant of cells lacking exopolysaccharide (Kobayashi and Iwano, 2012). At this point in time, the composition of the exopolysaccharide remains unclear. There are contradictory findings with regards to what the identity of the major polysaccharide found in the EPS is (Jones et al., 2014; Roux et al., 2015). Interestingly, the mixing of an *eps* mutant with a *tasA* mutant can restore biofilm structure (Branda et al., 2006; Dragos et al., 2017). This indicates that EPS and TasA are communal goods shared within the biofilm and is consistent with a division of labour at the level of gene expression. This co-operative behaviour was found to be an evolutionary stable relationship in the laboratory environment (Dragoš et al., 2018).

Extracellular DNA (eDNA) has been reported to be important for the biofilms of a number of strains. The release of eDNA by *B. subtilis* NCIB3610, grown as planktonic cultures in MSgg medium, has been reported (Zafra et al., 2012). Furthermore, the role of eDNA in biofilm formation has been studied to some extent in related *Bacillus* species (Randrianjatovo-Gbalou et al., 2017; Vilain et al., 2009), however, the role of eDNA in *B. subtilis* biofilm formation remains unknown.

γ-(D,L)-Polyglutamic acid (PGA) is an exopolymer which may contribute to biofilm formation in *B. subtilis* (Stanley and Lazazzera, 2005). The expression of the biosynthetic gene cluster that results in the production of PGA in a domesticated JH642 strain of *B. subtilis* enhanced adhesion to polyvinylchloride in a crystal violet assay (Stanley and Lazazzera, 2005). However, adhesion did

not seem to be perturbed when PGA production was blocked by the deletion of the *ywsC* PGA biosynthesis gene in an undomesticated strain (Stanley and Lazazzera, 2005). Additionally, PGA was not found in a chemical analysis of the extracellular matrix of the model biofilm forming strain NCIB3610 and the deletion of *ywsC* in the wild-type *B. subtilis* isolate RO-FF-1 did not lead to a decrease in biofilm formation (Morikawa et al., 2006). Despite this, PGA was found to be a major component of the extracellular matrix of the environmental isolate *B. subtilis* B1 (Morikawa et al., 2006). As a side note, hyper-production of PGA by *Bacillus subtilis var. natto* is responsible for the stringy texture of the fermented Japanese food natto (Shih and Van, 2001). Furthermore, the PGA capsule of the pathogenic species *B. anthracis* is responsible for evasion from the immune system (Jang et al., 2011).

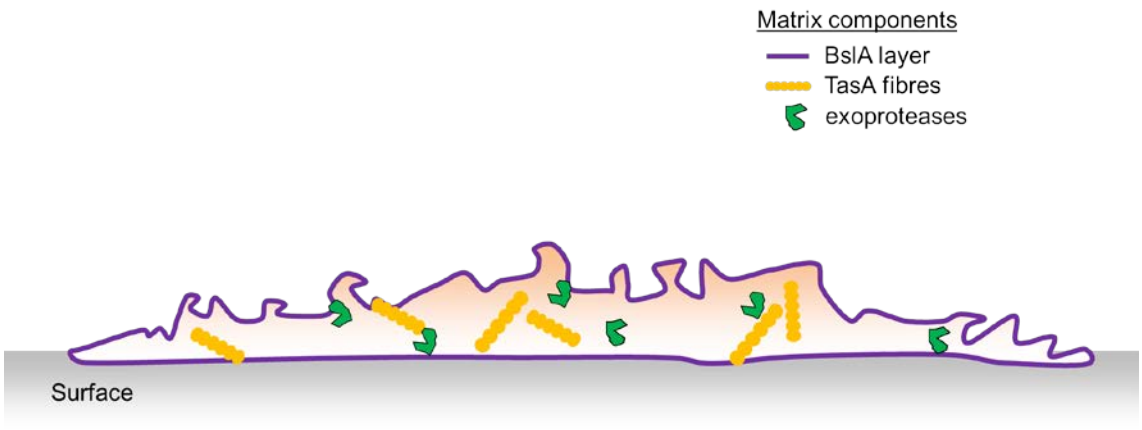


Figure 7: The localisation of matrix components within the *B. subtilis* biofilm. BslA (purple) forms a hydrophobic layer localised to the surface on the periphery of the biofilm. TasA fibres (yellow) are thought to extend from the surface of biofilm embedded cells, it is not clear whether TasA fibres are found cell-free within the matrix. *B. subtilis* is a prolific secretor of exo-proteases (green) and studies have shown that secreted-proteases are expressed in a spatio-temporal manner within the biofilm. Exopolysaccharide is essential for wild-type biofilm formation but is not clear what the composition of EPS is and how it is localised within the biofilm. Finally, NCIB3610 has been reported to produce eDNA but it is not clear whether it is important for biofilm formation. TapA is a secreted protein and it has been reported to be needed for TasA fibre formation but more recent work has shown that TasA forms fibres

spontaneously *in vitro* without the need of TapA. The current model of TapA function places it in association with the cell wall of *B. subtilis* cells. This figure is adapted from Hobley et al., 2015.

1.7. TasA fibres are a major component of the *B. subtilis* biofilm

The products of the *tapA-sipW-tasA* operon interact to generate the protein fibre component of the *B. subtilis* matrix (Branda et al., 2006; Romero et al., 2010). The model proposed for the formation of these fibres is that TasA and TapA interact to form amyloid fibres which are anchored to the surface of the bacterial cell by TapA (Romero et al., 2014). It is for this reason that the name TapA, for TasA anchoring/assembly protein A, was given to the protein formerly known as YqxM. TasA is the major component of these fibres and the deletion of *tasA* blocks the formation of complex, rugose biofilms *in vitro* and on plant roots (Beauregard et al., 2013; Branda et al., 2006). In this section I am going to summarise the research that has led to the current understanding of the function of TasA and TapA.

1.7.1. The secretion of TasA and TapA is dependent on SipW

The *tasA* gene encodes a 263 amino acid protein with a sec-dependent signal peptide (Stöver and Driks, 1999c). Therefore, the mature, secreted protein is estimated to have a molecular weight of 25.7 kDa after cleavage of the signal peptide by the signal peptidase SipW (Stöver and Driks, 1999b). The *tapA* gene encodes a 253 amino acid protein predicted to have a molecular weight (MW) of 29.08 kilo Daltons (kDa). The secreted TapA protein is predicted to be 24.18 kDa in MW due to the removal of the signal peptide by a signal peptidase, which in the literature is reported to be SipW (Stöver and Driks, 1999a). SipW has been shown to have a bifunctional role in biofilm formation in *B. subtilis* JH642 (wild-type isolate) as disruption of the catalytic activity of SipW did not affect the ability of a biofilm to form on a solid surface (Terra et al., 2012). The

other role of SipW is thought to be in the upregulation in the expression of genes encoding matrix components (Terra et al., 2012).

1.7.2. TasA as a fibrous protein

Originally, TasA was described as a spore-associated protein with antibacterial activity (Stöver and Driks, 1999b). It is now known that TasA also forms long fibres which are essential for the integrity and structure of biofilms (Branda et al., 2006) and that TasA has an additional role in sliding motility (van Gestel et al., 2015). I am going to briefly summarise what is currently known about the role of TasA in biofilm formation. Using immunoblotting, TasA was detected in the matrix of *B. subtilis* NCIB3610 cells grown as pellicles (Branda et al., 2006). Next, by using gold labelled anti-TasA antibodies, in combination with transmission electron microscopy, it was found that TasA forms fibres that emanate from the surface of the cell (Romero et al., 2010). The authors observed the TasA fibres after 24 h of growth as pellicles at 30^{o}C on MSgg media, the fibres were approximately 10-25 nm in width and variable in length. Additionally, they found that purified TasA fibres could be added exogenously to a Δ*tasA* mutant strain to restore biofilm formation in floating pellicles, but that the TasA fibres could not rescue a Δ*tapA* mutant biofilm defect (Romero et al., 2011). It is important to note that there are many components in the *B. subtilis* biofilm and that these components in combination allow for the development of complex architecture. Consistent with this, when the *bslA* (formerly known as *yuaB*) gene was knocked out the over-expression of the *tasA* and *eps* genes could not rescue the biofilm phenotype (Patrick and Kearns, 2008).

1.7.3. A high resolution structure of TasA

The first X-ray crystal structure of the TasA protein was published in late 2017 (Diehl et al., 2018). $TasA_{28-239}$, with selenomethionine incorporated, was crystallized as a monomeric protein at pH 4.6 and a high resolution structure was obtained (**Figure 9**A). The structure of the protein is described as a jellyroll fold with 2 antiparallel β-sheets with 6 small helices and extended loop regions. By analysing the rigidity of the structure it was concluded that there is a "dynamic section" which is proposed to convey structural flexibility that facilitates a conformational change suggested to occur upon fibre formation (Diehl et al., 2018).

1.7.4. TapA anchors TasA fibres to the surface of the cell

In 2004, TapA was implicated in biofilm formation for the first time (Hamon et al., 2004). It was found that, in the strain background of *B. subtilis* JH642, the disruption of the *tapA* gene led to reduced levels of biofilm formation on a plastic surface in a microtitre plate assay (Hamon et al., 2004). Subsequent work found that deletion of *tapA* eliminated complex, rugose colony formation in the strain NCIB3610 (Chu et al., 2006). This led to further investigation into the functional role of TapA. Using immunocytochemistry TapA localization could be visualized as puncta located on the surface of cells isolated from pellicles with the aid of fluorescein isothiocyanate conjugated (FITC)- secondary antibodies (Romero et al., 2011). Using immunogold electron microscopic analysis it was found that TapA could be detected in association with the cell wall of cells grown as pellicles (Romero et al., 2011). Furthermore, using the same technique it was found that TapA associates with peptidoglycan sacculi prepared from biofilm grown *B. subtilis*. It was further demonstrated that a Δ*tapA* mutant displays fewer and altered fibres emanating from the surfaces of

cells (Romero et al., 2011). TapA has been shown to be detected by α-TapA immunoblot in cell and matrix fractions of processed biofilms as a 28 kDa band and a smaller processed band of around 24 kDa at 24 hours. The larger band is proposed to be the result of cleavage of the signal peptide. At 48 h only the 24 kDa band was observed and only in the matrix fraction (Romero et al., 2011). It is suggested that this is due to further processing to generate the smaller TapA form. The authors found by immuno-labelling TapA and TasA with antibodies attached to gold nanoparticles of distinct sizes that both proteins were found in the fibres extending from the surface of the biofilm cells in a ratio of TapA 1:10 TasA. In a *tapA* mutant the levels of TasA protein are reduced as determined by immunoblot, meaning that a *tapA* mutant is broadly analogous to a Δ*tapA ΔtasA* double-mutant (Romero et al., 2011). Finally, it was found that Δ*tapA* and Δ*tasA* mutants do not generate wild-type biofilms upon mixing of the 2 strains (Romero et al., 2011). This indicated that TapA is not a shared resource within the biofilm in the way that TasA has shown to be (Dragoš et al., 2018). Given the above information the current model for TasA fibre formation is shown in **Figure 8**.

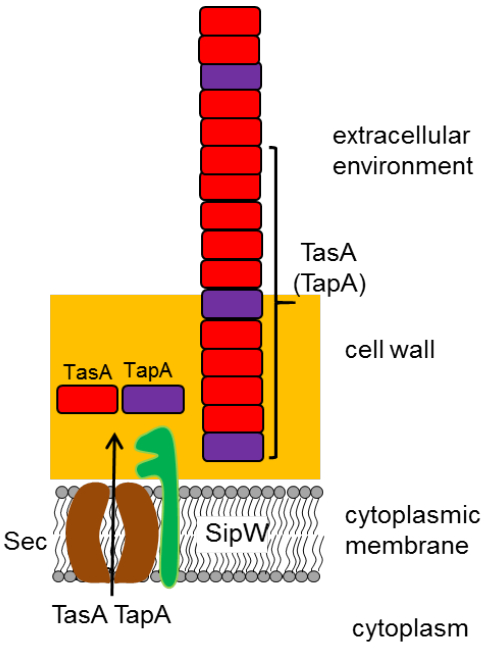


Figure 8: A model for TasA fibre formation. *B. subtilis* TasA fibres are reported to be cell-wall anchored and are essential for wild-type biofilm formation. TasA and TapA are reported to be transported by the Sec system with SipW as the signal peptidase that cleaves the signal peptide of both TasA and TapA. TapA is thought to anchor TasA to the cell wall making up a small proportion of the monomers composing TasA fibres. The biogenesis of TasA fibres is dependent on the products of the *tapA-sipW-tasA* operon. This figure is adaption from a publication (Hobley et al., 2015).

1.7.5. *In silico* 3D structure analysis for TapA predicts a β-sheet based structure

As things stand there is limited experimental insight into the structure of TapA and no x-ray crystal structure currently available. For this reason I used the online server RaptorX to predict the tertiary structure of TapA (Källberg et al., 2012). The result of this was the prediction of an α-helix from amino acids 8-33 which is indicative of a signal peptide region in the N-terminus of the protein (**Figure 9**B). The region between amino acids 72 and 191 forms a jellyroll-like fold which is similar to the β-sheet structure that has been reported for the solved TasA crystal structure (**Figure 9**B). Using alternative software, in the Phyre^2 server to predict the TapA structure resulted in a somewhat similar prediction (not shown). In this case an α-helix was predicted for amino acids 3-

37 with a TasA-like β-rich structure from amino acids 81-159 consisting of 6 β-strands (Kelley et al., 2015). Additionally, there were 5 α-helices on the C-terminal side of the β-rich region which are of variable lengths. Both Phyre[2] and RaptorX predicted structures contained extensive disordered regions.

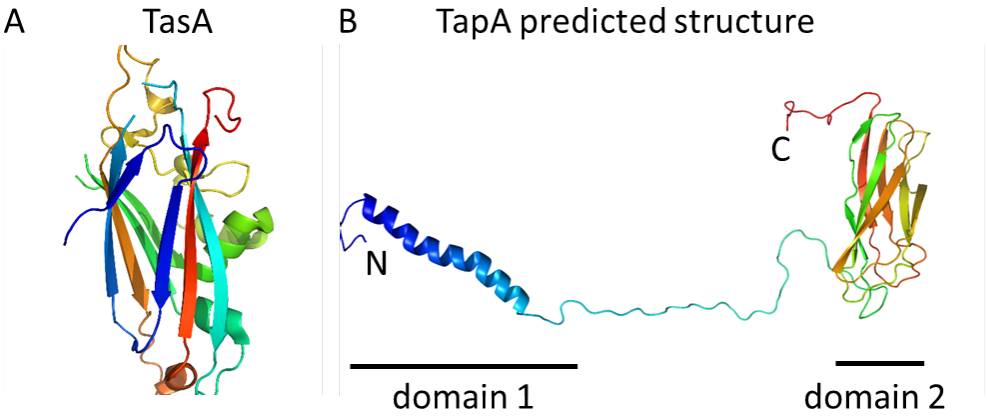


Figure 9: X-ray crystallography was used to solve a TasA structure of high-resolution. (A) The solved crystal structure of a $TasA_{28-239}$ monomer showing a β-strand rich structure with a consisting of 2 β-sheets to produce a jellyroll fold. **(B)** The predicted tertiary structure of full-length TapA as determined using the RaptorX server. The model is informed by the alignment of the TapA sequence with the protein sequence of the following known structures: the α-helix shown in blue (domain 1: amino acids 1-71) is informed by alignment with the structure 2a65:a (Yamashita et al., 2005) (p-value: $2.76^{-}02$, score: 26, uGDT(GDT):27(38)). The prediction of an α-helix is consistent with the presence of an N-terminal signal sequence. The structure, 2KL6A (unpublished), aligns to the β-sheet structure (domain 2: amino acids 72-191). This takes the form of 2, 4 strand β-sheets forming a jellyroll-like structure resembling a β-barrel (p-value: $4.48^{-}03$, score 40, uGDT(GDT):50(42)). Note that the probability scores are low for the prediction of both domains.

1.7.6. TapA is thought to be cell wall-associated

It has been published that TapA is found in the cell wall fraction of cells grown as colonies or pellicles which may be how TapA is anchored to the cell surface (Romero et al., 2014). It was initially reported that the addition of the D-amino acids as a mixture of D-leucine, D-methionine, D-tyrosine, and D-tryptophan could disrupt biofilm formation (Kolodkin-Gal et al., 2010). It was suggested that

this involved the incorporation of D-amino acids in the peptide side chains of the peptidoglycan making up the cell wall. It was further demonstrated that *B. subtilis* isolates which had developed a 3' mutation in the *tapA* gene were resistant to the biofilm disassembly effects of D-amino acids. Importantly, follow-up work from the same laboratory found that there was not a biofilm specific effect of D-amino acids and that they were functioning to have a toxic effect on protein synthesis in *Bacilus subtilis,* impacting on overall growth (Leiman et al., 2013; Leiman et al., 2015; Sarkar and Pires, 2015).

1.8. Amyloid fibres

Congo red and Thioflavin T (ThT) dyes have been used for many years to stain and detect amyloid proteins associated with human disease (Bennhold, 1922; Divry, 1927; Saeed and Fine, 1967). It has been observed that the deletion of *tasA* or *tapA* reduces the binding of Congo red to *B. subtilis* biofilms (Romero et al., 2010; Romero et al., 2011). TasA fibres have been reported to bind Congo Red and ThT, and for this reason TasA fibres purified *ex vivo* have been classified as functional amyloid fibres (Romero et al., 2010). I will first describe the properties associated with amyloid fibres and using this information provide evidence that TasA fibres do not display the characteristics of an amyloidogenic fibre.

1.8.1. What are amyloid proteins?

The aggregation of amyloid proteins in the brain is a shared hallmark of Alzheimer's, Parkinson's and Huntington's disease (Selkoe and Hardy, 2016). Amyloid fibres are highly stable forming from monomers which nucleate into fibrillar deposits displaying a characteristic cross-β-sheet structure when analysed by X-ray diffraction such as for α-Synuclein (**Figure 10**) (Breydo et al.,

2012). Amyloid fibres have a striking structure with β-sheets arranged perpendicular to the axis of the fibres.

In recent years attention has been drawn to “functional” amyloid fibres present in the biofilms of bacterial species such as the Curli fibres of *E. coli*. There is debate as to whether “amyloid” is an appropriate descriptor for these proteins as it is thought that amyloid fibrils can be induced in most proteins given the right solvent and other conditions such as pH or temperature (MacPhee and Dobson, 2000). Therefore, it has been postulated that amyloid fibril formation is a ‘generic’ property of the polypeptide backbone possible given the appropriate solution conditions (Lamour et al., 2017).

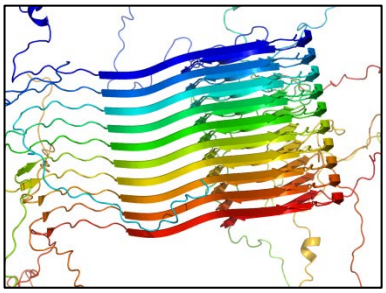


Figure 10: The structure of α-sunclein is made up of parallel, in register β-sheets. The solid-state NMR (nuclear magnetic resonance) of α-synuclein demonstrates this characteristic conformation displayed by amyloid fibrils (PDB:2N0A) (Tuttle et al., 2016).

1.8.2. TasA does not form amyloid fibres

It was found that the polymerisation of TasA could be measured, in a matrix preparation extracted from wild-type cells, by the addition of the amyloid-binding dye thioflavinT (ThT) (Romero et al., 2014). In this case, increased fluorescence over time was indicative of fibre formation from smaller oligomers (Romero et al., 2014). The polymerisation of the TasA fibres was increased when recombinant TapA protein was added to the preparation, as a control it was demonstrated that TapA did not bind ThT (Romero et al., 2014).

However, in contradiction to the work described above we have recently published that the Congo red and ThT binding activity displayed by Δ*tasA*-derived protein extracts and wild-type extracts was similar (Erskine et al., 2018). It should be noted that the specificity of Congo Red and ThT is not absolute. ThT has been shown to display enhanced fluorescence in the presence of amorphous aggregates of bovine serum albumin and lysozyme, for example, (Yang et al., 2015) and Congo red is also non-specific (Howie and Brewer, 2009). In other words binding of dyes does not confirm amyloid structure but is merely suggestive.

Given this word of caution, it was found upon further analysis that TasA fibres lacked the “cross-β“ structure displayed by amyloid-like fibrils (Erskine et al., 2018). Furthermore, monomeric TasA was found to have the same secondary structure as the fibre form (Erskine et al., 2018). This indicated that TasA fibres are instead a linear assembly of monomers, with no profound structural rearrangement occurring. This is in conflict with other reports which state that TasA undergoes a conformational change during the process of biofilm formation (Diehl et al., 2018). Another hallmark of amyloid-fibres is insolubility in SDS and a resistance to degradation by proteases. It was found that TasA was soluble in SDS, however, the fibre form was resistant to degradation by proteases (Diehl et al., 2018; Erskine et al., 2018).

In conclusion, although TasA fibres display some of the characteristics of amyloid fibres (staining with amyloid dyes and resistance to proteolysis), the evidence suggests that TasA fibrils are assembled from monomers with a similar structure to the globular subunits of the fibre (Erskine et al., 2018).

1.9. Summary

In summary, the study of *B. subtilis* biofilms has elucidated the incredible diversity in specialisation of cells within the biofilm. From matrix production, spore formation, motile cells, competent cells and surfactin producing cells. This demonstrates the complexity within the bacterial biofilm. The structure of the biofilm is conveyed by the embedding of cells in a self-produced extracellular polymeric matrix. The properties and function of the matrix have become a focussed area of interest in recent years with components such as BslA having exciting properties with potential biotechnological applications. There is now a field of research looking into fibrous biofilm proteins and their physical properties including those classified as functional amyloid fibres.

1.9.1. Key aims and objectives

In the current work biochemical, bioinformatic and genetic approaches are employed to aid in the understanding of how the biofilm matrix is assembled and the role of the secreted matrix protein TapA in this process. The aim of the first results section (3.1) is to analyse the active form of TapA within the biofilm. It is hypothesised that TapA functions as a secreted protein where the mature form of the protein consists of amino acids 44-253. Additionally, *in vitro* analysis of the purified TapA protein is undertaken in an attempt to understand the functional form *in vivo*. The second results section (3.2) documents our attempts to develop a model to probe TapA and TasA interaction. The aim of this section was to devise a system in which the molecular determinants of TapA/TasA interaction could be elucidated. Finally, in section 3.3 we carried out a genetic analysis with the aim to determine the minimal functional genetic unit of *tapA*. Using this insight we attempted to find the functional form of TapA *in vivo*.

2. Methods

2.1. Media and antibiotics

2.1.1. LB and antibiotics

Lysogeny broth (LB) was prepared using 10 g typtone, 10 g NaCl and 5 g yeast extract for 1 litre. LB agar was prepared by solidifying with 15 g of select agar (Sigma; A5054) for growth of *B. subtilis* and *E. coli*. For antibiotic selection with *B. subtilis* strains antibiotics were used at the following final concentrations: erythromycin (0.5 µg/mL), spectinomycin 100 µg/mL and MLS, erythromycin (0.5 µg/mL) together with lincomycin (12.5 µg/mL). For antibiotic selection with *E. coli* then ampicillin was used at a concentration of 100 µg/ mL.

2.1.2. MSgg growth medium

To make Minimal Salts glycerol glutamate (MSgg) 5 mM potassium phosphate, 100 mM MOPS at pH 7.0 with agar to a final concentration of 1.5% was autoclaved and cooled to 55^{o}C before being supplemented with 2 mM $MgCl_2$, 700 µM $CaCl_2$, 50 µM $FeCl_3$, 50 µM $MnCl_2$, 1 µM $ZnCl_2$, 2 µM thiamine, 0.5% (v/v) glycerol, 0.5% (w/v) glutamic acid. For induction of gene expression from the P_{spank} promoter Isopropyl β-D-1-thiogalactopyranoside (IPTG) was included at a final concentration of between 25 µM to 1 mM. For MSgg broth the same recipe was followed without the addition of agar.

2.1.3. Phage media

The growth medium used to propagate and transfer SSP1 phage was prepared by adding $MgSO_4$ and $MnSO_4$ to lysogeny broth (LB), to a concentration of 10 mM and 1 µM respectively. Phage agar was made by adding agar to a concentration of 1.5% (w/v) and to 0.7% (w/v) for phage top agar. To prevent cell lysis by lysogeny, the cells plus phage mixtures were plated onto phage agar containing sodium citrate at a concentration of 10 mM.

2.1.4. Modified competence (MC) media

10x MC media was adapted from Konkol *et al* (Konkol et al., 2013) and prepared as a solution containing 10.7 g K_2HPO_4, 5.2 g KH_2PO_4, 20 g dextrose, 0.88 g sodium citrate dehydrate, 2.2 g l-glutamic acid monopotassium salt, , 1 g casein enzymatic hydrolysate per 100 mL made up with milliQ-H_2O. 3 µL of 1M $MgSO_4$ and 1.75 µL of 500 mM $FeCl_3$ was added fresh per 1 mL of 1x MC media prepared. A single colony was inoculated into 2 mL of MC broth and grown for 4.5h at 37^oC with shaking, after which 400 µL of culture was mixed with 500 ng of plasmid, 1 µg of PCR product or approximately 1 µg of genomic DNA (gDNA) and grown for a further 1.5 h at 37^oC with shaking. Finally, the culture was plated onto LB plates with the required antibiotic for selection.

2.2. General overview of cloning and strain construction

Escherichia coli strain MC1061 [F' *lacIQ lacZM15 Tn*10 (*tet*)] was used for the maintenance and stocking of plasmids generated by restriction endonuclease cloning. Generally, this involved the cloning of genetic material, such as a variant of the *tapA* gene into the vector pDR110 which allowed for the integration, by homologous recombination, of the variant into the *amyE* locus of the *B. subtilis* genome. As NCIB3610 is not naturally competent, gene variants were first introduced into *B. subtilis* strain 168, which is naturally competent, by transformation. The genetic information could then be transferred to the biofilm forming strain NCIB3610 using SPP1 phage-mediated transduction. All experiments concerned with biofilm formation and TapA function were either carried out in this strain or in 3610 $comI_{Q12L}$ which is a modified, competent version of the parental strain NCIB 3610. For a full list of strains generated in this study please see strain table .

2.2.1. Plasmid extraction

Plasmids were isolated from *E. coli* MC1061 grown as a 5 mL LB overnight culture with the presence of antibiotic selection. Briefly, the culture was pelleted by centrifugation, for 10 min at 3,800 x g, and the plasmid was extracted from the cells using the QIAprep® Spin Miniprep Kit from Qiagen following the manufacturer's instructions. The use of the PB buffer is essential given the host strain genotype.

2.2.2. Polymerase chain reaction (PCR)

For PCR amplification (Saiki et al., 1985) of inserts to be cloned into the host vector pDR110 the commercially available Q5® High-Fidelity 2X Master Mix from New England Biolabs (NEB) was used. Reactions contained the following: 200 µM dNTPs, 2.0 mM Mg^{2+}, forward and reverse primers at a concentration of 0.5 µM and genomic DNA or plasmid template at between 50-100 ng. The total volume of the reaction, made up to with ddH_2O, was typically between 30 and 50 µL. The reaction was carried out in a thermocycler, see **Table 1** for more detail in regards to the cycling conditions. PCR amplification products were analysed by agarose gel electrophoresis before being gel extracted to remove enzymes using the QIAquick Gel Extraction Kit from Qiagen following the manufacturer's instructions.

Step	Temperature	Time
Initial denaturation	98^{o}C	30 sec
Denaturation	98^{o}C	10 sec
Annealing	50-72^{o}C (primer T_m-dependent)	30 sec
Extension	72^{o}C	30 sec per kb of DNA
Final extension	72^{o}C	2 min
Hold	4-10^{o}C	

Table 1: The PCR cycling conditions used in conjunction with the Q5® High-Fidelity 2X Master Mix from NEB. The Denaturation to Extension steps were generally repeated for a total of 30 cycles.

2.2.3. Restriction endonuclease digest, phosphatase treatment and gel extraction

Restriction endonuclease digests generally involved the incubation of the full volume of the PCR amplification product with 5 µL of the recommended 10x buffer and 2 µL of each enzyme, usually SalI-HF and SphI-HF, obtained from NEB. The reaction was made up to a total volume of 50 µL using ddH_2O and incubated at 37^{o}C for 3-4 hours. To remove endonuclease enzymes the digestion reaction was purified using the QIAquick PCR Purification Kit from Qiagen following the manufacturer's instructions. Digests of plasmid vectors, for the hosting of amplified inserts, were carried out as above but with 1 µg of plasmid digested. Additionally, digested vectors were subjected to an alkaline phosphatase treatment to prevent vector re-ligation and increase efficiency of ligation. After the restriction digest was carried out 4 µL of Antarctic phosphatase (NEB) was added to the reaction along with 6 µL of the recommended 10 x buffer and incubated at 37^{o}C for 30 min, the enzyme was then heat-inactivated at 80^{o}C for 2 min. Digested vectors were analysed by agarose gel electrophoresis to confirm digestion, followed by gel extraction to purify the vector and remove enzymes (carried out using the QIAquick Gel Extraction Kit from Qiagen).

2.2.4. Agarose gel electrophoresis

Following all enzymatic reactions (for example PCR products or restriction endonuclease digests), DNA products were resolved by horizontal agarose gel electrophoresis. For this 0.8 - 2% (w/v) agarose gels with 1X GelRed nucleic acid stain (Biotium) were prepared in 1X Tris base, acetic acid and ethylenediaminetetraacetic acid (EDTA) (TAE) buffer (see Table 6: reagents). DNA samples were prepared with 6X gel loading dye (NEB) and 3 µl of

HyperLadder™ 1kb DNA ladder (from BioLine) was loaded alongside the samples. Electrophoresis was performed in 1X TAE buffer under constant voltage at 120 V for 60 minutes. Nucleic acids were imaged using the Gel Doc XR system (BioRad).

2.2.5. DNA Ligation

Purified DNA fragments were ligated into linearised plasmids using T4 DNA ligase (NEB). Reactions were prepared as follows: 400 U T4 DNA ligase, 1X T4 DNA ligase buffer with varying volumes of plasmid and insert (normally the reaction was based on a vector to insert volume ratio of 1:3), with ddH_2O added to give a total reaction volume of 20 µl. Reactions were incubated overnight at room temperature and 10 µL of the reaction was transformed into competent *E. coli* MC1061 cells. The *E. coli* cells were grown at 37^oC, overnight, in the presence of antibiotic selection specific for the resistance conferred by the vector, typically ampicillin for the pDR110 and pMAD cloning vectors.

2.2.6. Site-directed mutagenesis

On occasions where a variant of the *tapA* gene was desired to have a change in the 3'-end of the *tapA* coding region, then this mutation, depending on the proximity to the end of the sequence, usually within 20 bp, could be introduced by varying the sequence of the reverse primer in a standard PCR reaction using Q5 polymerase. For all other site-directed mutations then a QuickChange protocol was carried out using KOD Hot Start DNA Polymerase (Novagen). Briefly, forward and reverse primers were designed to overlap with the DNA template at the same position. The primer sequence varied so that they contained the desired DNA sequence. For example, the primer sequence could be CUU (encoding Leucine) instead of GCU (Alanine) found at that position in

the DNA template, to create an amplicon which encoded CUU instead of GCU. Single reactions were prepared to contain the following components: 0.2 mM dNTPs, 1 mM MgSO, 0.5 µM forward primer, 0.5 µM reverse primer, 100 ng template, 1 U KOD Hotstart polymerase in PCR buffer diluted to 1X in a total volume of 50 µl. Cycling conditions are shown in **Table 2**.

Step	**Temperature**	**Time**
Initial denaturation	95^{o}C	2 min
Denaturation	95^{o}C	20 sec
Annealing	57^{o}C (primer T_m-dependent)	10 sec
Extension	70^{o}C	4 min (30 sec per kb of DNA)
Final extension	70^{o}C	5 min
Hold	4-10^{o}C	

Table 2: The PCR cycling conditions used in conjunction with the KOD Hot Start DNA Polymerase. The Denaturation to Extension steps were generally repeated for a total of 20 cycles.

2.2.7. Colony PCR and sequencing

Colony PCR was used to analyse transformants to determine whether the competent *E. coli* had successfully been transformed with the ligated insert-vector construct. A single colony was suspended in 12 µL of ddH_2O, 3 µL of this mixture was spotted onto a plate and a further 3 µL was used as the template in a PCR reaction. The PCR reaction was carried out using 10 µL of GoTaq®G2 Green Master Mix 2x, 3 µL of template DNA, 0.5 µM of forward and reverse primers and made up to a final volume of 20 µL with ddH_2O. PCR products were analysed by agarose gel electrophoresis. Positive PCRs resulted in the corresponding clone being stocked. The clone was confirmed by sequencing. Briefly, plasmid DNA was extracted from an overnight 5 mL culture of the clone. The plasmid was then sequenced with primers 39F and 39R (see primers: Table 9). All sequencing was performed by the DNA Sequencing and Services at the University of Dundee (http://www.dnaseq.co.uk). Raw nucleotide

sequencing data was analysed using the CLC Main Workbench version 6. BLAST was used to compare query nucleotide sequences by pairwise alignment to the expected sequence (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.2.8. Stocking *E. coli* strains

E. coli strains containing plasmids constructed by cloning were stocked as follows: 5 mL LB broth with antibiotic (usually, if not exclusively, ampicillin) was inoculated with a single colony and grown overnight for 16 h after which 1 mL of the culture was mixed with 800 µL of 50% (v/v) glycerol, giving a final glycerol concentration of 22% (v/v), and stored in a cryovial at -80^{o}C.

2.2.9. *B. subtilis* genomic DNA extraction

A single colony was used to inoculate 3 mL of LB broth and the culture was grown at 37^{o}C with shaking for 3 h or alternatively overnight to increase yield. Bacterial cells were then pelleted by centrifugation and the pellet was re-suspended in 180 µL of enzymatic lysis buffer and incubated at 37^{o}C for 30 min. Genomic DNA was extracted using the Qiagen DNEasy Blood and Tissue kit following the manufacturer's instructions.

2.2.10. Generating competent *B. subtilis* 168 and transformation

Typically for this project, once a plasmid was constructed *in vitro* and replicated in the *E. coli* host strain as described above it was then transformed into the genetically competent strain *B. subtilis* 168. Firstly, to generate competent cells, strain 168 was plated out from a frozen stock and grown overnight at 37^{o}C. A single colony was used to inoculate 100 µL of LB broth and this was plated onto an LB plate which was grown at 22^{o}C overnight. The next day the lawn was washed with 5 mL LB broth to re-suspend the cells and the OD_{600} of the resulting solution was measured. The solution was used to inoculate 12.5 mL of

SpC media in a 250 mL conical flask to a starting OD_{600} of 0.01. The culture was grown at 37°C with shaking until 2 hours past the onset of stationary phase. The culture was then diluted 1:10 in pre-warmed SpII and grown for a further 90 minutes at 37°C with aeration. The cells were then collected by centrifugation and re-suspended in 1/10th the volume of supernatant. They were either used directly for transformation or stored at -80°C for later use with the addition of glycerol to the cells to give a final concentration of 10% (v/v) glycerol. Transformation was carried out by mixing 100 µL of cells 1:1 with transformation buffer with plasmid DNA, typically 500 ng, and this mixture was incubated at 37°C for 20 minutes before plating out on LB agar with antibiotic selection. Selection was usually on spectinomycin in the case of transformations with plasmid pDR110. Screening of colonies for successful integration at the *amyE* locus was carried out by patching the colonies onto LB plates supplemented with 1% (w/v) soluble starch. On addition of Lugol's iodine solution (Sigma) the starch containing agar turns blue/black. *B. subtilis* cells with an intact *amyE* gene can break down the starch to form a halo in the presence of the iodine solution. Patches were selected that were negative for amylase and stocked.

2.3. Integration of genes onto the *B. subtilis* chromosome and chromosomal deletions

2.3.1. Integration of genetic variants at the *amyE* locus

The donor 168 strain was grown at 37°C with shaking overnight in TY broth and 200 µL of this culture was mixed with the same volume of phage broth and 1 µL of SPP1 phage and incubated at 37°C for 15 minutes. The mixture was then mixed with 3 mL of molten phage agar which was poured on top of a phage

agar plate. The plate was placed at 37^{o}C until plaques had formed in most of the area of the top agar (approximately 3-5 hours). The top agar was then harvested using a sterile loop and re-suspended in 5 mL of phage broth and vortexed for 1 min to break up the agar. The phage were then eluted using a syringe with a 0.2 μM filter to isolate the phage from the donor strain cells. The recipient strain was grown in TY broth for 3-5 h before 900 μL of this culture was diluted with 9 mL of phage broth to which 60 μL of phage was added and this was incubated for 30 min at 37^{o}C. The cells were finally plated onto phage agar with sodium citrate and antibiotic for selection. To increase the efficiency of transduction, when necessary, the volume of phage was varied. Screening on LB agar plates with 1% (w/v) was carried out as before to identify colonies that had undergone successful integration of the genetic construct at the *amyE* locus.

2.3.2. In-frame markerless deletion strains of *B. subtilis* NCIB3610

The Δ*tapA* deletion strain was generated by allelic exchange in a method similar to that previously published using the pMAD plasmid (Arnaud et al., 2004). Briefly, a 395 bp upstream region was amplified by PCR with primers NSW1308 and NSW1332 and a 641 bp downstream region was amplified using primers NSW1333 and NSW1334. Both fragments were cloned into the pMAD (pMinimad) vector to generate plasmid pNW685 containg an in-frame deletion construct. As NCIB3610 is not naturally competent the plasmid was first transformed into *B. subtilis 168.* Briefly, the plasmid was incubated with competent 168 cells and transformation buffer at a volume ratio of 1:1, for 1 h at 30^{o}C, as that is the permissive temperature for replication of the plasmid which has a temperature sensitive origin of replication. The mixture was then plated

onto selective LB MLS plates. The pMAD plasmid confers resistance to erythromycin and lincomycin which are of the macrolide-lincosamide-streptogramin (MLS) group. Several transformants were pulled together and inoculated into 5 mL TY broth with MLS. The culture was then incubated at 30^{o}C with aeration for 16 h. Phage carrying the constructed pMAD plasmid were generated from this culture as before and *B. subtilis* NCIB 3610 was transduced with this phage in an incubation step for 1 h in TY media at 30^{o}C. Successfully transduced colonies were selected by growth for 48 h on TY agar plates supplemented with 10 mM sodium citrate and MLS. Sodium citrate was included to stop phage activity and MLS acted as selection for strains carrying the transduced plasmid. Several transduced colonies were pooled together to inoculate 5 mL of LB broth supplemented with MLS and grown at 37^{o}C for 16 h with shaking. Growth at the non-permissive temperature and in the presence of MLS antibiotics selects for the integration of the plasmid including the resistance cassette into the genome. Cells that had undergone integration were selected for by plating the culture onto LB MLS agar. Several post-integration colonies were pooled together and inoculated into 5 mL of LB medium. The culture was grown under static conditions for 16 h at 30^{o}C at which point the culture was incubated with aeration for a further 4 h at 30^{o}C. Growth at the lower temperature allows for the loss of the plasmid from the chromosome. Colonies were plated onto LB and once grown replica patched onto LB and LB MLS. MLS-sensitive colonies were then screened by colony PCR and those which gave a positive signal were then sequenced to confirm the scar left after deletion of the native gene and the integrity of the regions upstream and downstream of the gene. The Δ*tapA* mutant was generated using the above method by Laura Hobley. The strain Δ*tapA* Δ*tasA sinR* (NRS5749) was

generated as part of the current work, using the same method as above by deletion of the *tapA* gene but in the strain background of Δ*tasA sinR* (NRS5248).

2.3.3. Making exoprotease-free *B. subtilis* 3610

To make markerless deletions in all 7 secreted proteases from *B. subtilis* 3610 the BKE collection was utilised alongside a competent strain of *B. subtilis* 3610. The BKE collection is a knockout library of *B. subtilis* 168 containing over 4000 strains, with each strain containing a single gene deletion which has been replaced with an Erm casette (Genomic DNA was extracted from strains containing an Erm resistance cassette in place of the gene of interest and used to transform competent *B. subtilis* 3610 *comI* (Konkol et al., 2013) before selection on LB erythromycin plates. The erythromycin cassette contains *lox* sites and was subsequently removed leaving a 150 bp scar by the action of a Cre recombinase and this gene was expressed from plasmid pDR244. The plasmid is heat-sensitive and was then cured by growing at 42^{o}C. In cases in which transformation with genomic DNA proved unsuccessful then the mutation was introduced by phage transduction with SPP1 phage. Thanks to Tetyana Sukhodub for assistance in generating exoprotease-free 3610 *comI.*

2.3.4. Stocking of *Bacillus* strains

All *Bacillus* strains whether received from the Bacillus genetic stock centre (BGSC) or generated by phage transduction or transformation were twice passaged to generate single colonies by streaking onto LB agar plates (containing antibiotic where required) before a lawn plate was set up for stocking. To make a lawn plate a single colony was suspended in 100 µL of LB broth which was then plated onto LB agar and grown for 16 h at 22^{o}C. When

adequate growth was attained the lawn was washed with 3 mL of LB broth and 1 mL of the broth-cell mixture was mixed with 800 µL of 50% (v/v) glycerol, giving a final glycerol concentration of 22% (v/v), and stored in a cryovial at -80^{o}C.

2.4. Biofilm phenotype analysis

2.4.1. Biofilm formation assay

Biofilms were prepared by growing *B. subtilis* (and strains detailed in **Table 7**) in 3 mL of LB broth at 37^{o}C with aeration for 3-4 hours before 5-10 µL of culture was spotted onto an MSgg agar plate. The plate was then incubated at 30^{o}C for 48 hours. Biofilms were imaged using the a Leica MZ16 stereoscope (Leica Microsystems). For strains which had been generated to harbour gene variants integrated into the *amyE* locus of the genome under the expression of the P_{spank} promoter then IPTG was included in the MSgg agar to induce expression of the gene. Typically, a concentration of 25 µM IPTG was used to induce expression of *tapA* gene variants.

2.4.2. Exogenous assays using fTasA and TapA peptide ($TapA_{44-57}$)

A 3 mL LB broth culture was inoculated with a single colony and grown for 3-4 hours at 37^{o}C with shaking at 200 rpm. The culture was then centrifuged at 3220 x g for 10 minutes. Half of the supernatant volume (1.5 mL) was removed and the pellet re-suspended in the remainder of the supernatant, to provide a 2x concentrated culture solution. Then 10 µL of the culture solution was mixed 1:1 with 10 µL of fTasA protein or 10 µL of purification buffer (see Table 6) as a control. 10 uL of the cell/protein mixture, containing 20 µg of TasA, was then spotted onto MSgg plates. Biofilms were grown at 30^{o}C for 48 h. fTasA was

kindly supplied by Elliot Erskine and was purified under the conditions outlined in (Erskine et al., 2018).

The peptide $TapA_{44-57}$ was synthesised by CS Bio Co., California, USA and re-suspended at a concentration of ~10 µg/µL in 25 mM phosphate buffer (pH 7.0). This concentration is approximate as the scales were not sensitive to the amount of peptide being weighed. The exogeneous addition of $TapA_{44-57}$ was carried out as above, for fTasA, with the following modifications. The cell culture was concentrated 4x by re-suspending in 750 µL of culture supernatant to allow an increased amount of peptide to be added. 2 µL of the culture solution was mixed with 8 µL of peptide (80 µg total) before spotting the full 10 µL onto MSgg agar.

2.4.3. Protein purification

B. subtilis $TapA_{44-253aa}$ was separated from a Glutathione S-transferase-tag with a tobacco etch virus (TEV) protease-cleavage site using the pGEX-6P-1 system. The pGEX-6P-1 plasmid carrying the gene encoding the TapA variant was transformed into *E. coli* BL21 (DE3) pLysS. Once transformed the cells were grown overnight in 5 mL LB broth, the culture was then inoculated into auto-induction media (Studier, 2005) supplemented with ampicillin (100 µg/mL) at a ratio of 1:1000 (vol:vol) in a total volume of 1 litre. The cultures were incubated at 30^{o}C with shaking for approximately 6-7 h at which point the temperature was reduced to 18^{o}C for overnight incubation. The cell culture was pelleted by centrifugation for 45 min at 5020 g and re-suspended in 25 mL purification buffer (Tris 25 mM and NaCl 250 mM (pH 7.6) supplemented with Complete EDTA-free proteinase inhibitors mixture (Roche). Cell lysis was carried out by sonication at an amplitude of 20% for a total of 6 minutes.

Unlysed cells and cell debris were removed by centrifugation at 27,000 x g for 20 min. The cleared lysate was mixed with 750 µL (per litre of culture) Glutathione Sepharose 4B (GE Healthcare) and gently agitated at 4^{o}C for at least 3 h to allow binding of GST to the beads. The lysate/bead mixture was loaded onto a single-use, 25 mL gravity flow column (Bio-Rad). The beads were washed using 50 mL purification buffer, collected and incubated overnight at 4^{o}C with agitation in 25 mL of purification buffer supplemented with 1 mM DTT and 0.5 mg of TEV protease to release the protein from the GST-tag. The solution containing TapA, TEV protease, free-GST and the beads was loaded onto the gravity flow column. This removed the used beads which stay on the column. The flow-through was added to 250 µL Ni-nitrilotriacetic acid agarose (Qiagen) slurry to remove the TEV protease and 750 µL glutathione sepharose 4B to remove the free GST-tag. The mixture was incubated at 4^{o}C with agitation overnight, and then passed through a gravity flow column. The purified protein was concentrated using a Vivaspin 20 concentrator (with a MW cut-off of 5,000 or 10,000 Sartorius). The protein was then analysed by running 28-30 µg on SDS-PAGE.

2.4.4. Time course of TapA$_{44-253}$ stability

Purified TapA proteins (both TapA$_{44-253}$ and TapA$_{34-253}$) were stored at 4^{o}C in purification buffer. 28 µg of protein was sampled at intervals of 7 days, mixed with 4 x sample loading buffer (see Table 6), β-Me reducing agent and boiled at 100^{o}C for 10 min and then stored at -20^{o}C. After a number of samples were obtained then they were run on 14% SDS-PAGE and the gel was stained with InstantBlue for analysis. Images of gels were taken using the Gel Doc XR system (BioRad).

2.4.5. Protein extraction from biofilm prior to immunoblot analysis

Biofilms were isolated from MSgg plates with a sterile loop and re-suspended in 250 µL of BugBuster solution (Millipore; 71456) using a syringe with a 23 x 1 gauge needle until solubilized. The samples were then sonicated at an amplitude of 25% power for 5 seconds. Sonicated biofilms were next incubated at 26^{o}C for 20 min at 1,400 rpm before centrifugation at full speed for 10 min in a benchtop centrifuge. The liquid phase was taken and used for SDS-PAGE or stored at -20^{o}C for future use.

2.4.6. Quantification of protein concentration

The protein concentrations of purified protein, whole cell lysates and biofilm lysates were determined using the Nanodrop spectrophotometer ND-1000 at an absorbance of 280 nm.

2.4.7. Sodium Dodecyl Sulphate - Polyacrylamide Gel Electrophoresis (SDS - PAGE)

Proteins were separated by SDS-polyacrylamide gel electrophoresis based on the method previously published (Laemmli, 1970) using the Mini-protean system (BioRad). The stacking gel contained: 125 mM Tris-HCl (pH 6.8), 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.1% (v/v) TEMED and 6% (v/v) polyacrylamide. The resolving gel contained 375 mM Tris-HCl (pH 8.8), 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.1% (v/v) TEMED. Polyacrylamide was added to reach a final concentration of either 12 or 14% (v/v). The components of the SDS running buffer used are found in **Table 6**. For analysis of purified proteins approximately 30 µg of protein was mixed with 4x NuPAGE LDS sample buffer and 10x NuPAGE sample reducing agent, alternatively, samples were mixed with SDS-PAGE loading dye and β-mercaptoethanol was used as the reducing agent.

Samples were boiled at 100^{o}C for 10 min. Samples were run alongside the Precision plus protein dual color standard (Biorad) which was used as the molecular weight protein marker. The electrophoresis was carried out in SDS running buffer at 120 V while the samples migrated through the stacking gel and 220 V thereafter. To visualise proteins, gels were stained using InstantBlue (Expedeon) and visualised using the Gel Doc XR system (BioRad).

2.4.8. Immunoblotting

Samples to be analysed by immunoblotting were first separated by SDS-PAGE. The proteins were then transferred to hydrophobic polyvinylidene difluoride (PVDF; Immobilon-P from Millipore) membranes with a 0.45 µM pore size by electroblotting using the Mini Protean system (BioRad) in Tris glycine transfer buffer (see **Table 6**) at a constant current of 100 mA for 90 min. Membranes were blocked with a 5% (w/v) semi-skimmed dry milk solution in TBS-tween 0.2% (v/v) for at least 1 hour at room temperature or overnight at 4^{o}C. The membrane was then incubated with the primary antibody for 45 minutes if αTasA was used or incubated overnight at 4^{o}C where αTapA or αTapA serum were used (see **Table 3**). The membrane was then washed 3 times with TBS-tween 0.2% (v/v) to remove unbound primary antibody before incubation with the species-specific secondary HRP-conjugated antibody (Pierce) for 1 hour at room temperature in TBS-tween 0.2% (v/v). 3 times washes with TBS-tween 0.2% (v/v) were carried out as before and development was induced with Enhanced Chemi-Luminescence reagents (ECL; BioRad Clarity). Imaging was carried out using the GeneGnome (SynGene).

Protein antibody raised against	Species raised in	Dilution	Source
Primary antibodies			
TapA (*B. subtilis*)	Rabbit	1:2,500 (incubated overnight)	This work
TapA serum (*B. subtilis*)	Rabbit	1:2,500 (incubated overnight)	A. Driks (Loyola University)
TasA (*B. subtilis*)	Rabbit	1:25,000 for 45 min	(Ostrowski et al., 2011)
Secondary antibody			
αRabbit	Goat	1:5,000 incubated for 1 hour	(commercial: Pierce)

Table 3: Antibodies used in this study.

2.4.9. Supernatant protease activity assay

Lawn plates were set up by re-suspending a single colony in 100 µL of LB medium and plating the re-suspension onto LB agar. Lawn plates were incubated overnight at room temperature. After approximately 24 h lawns were re-suspended in 5 mL of MSgg broth using a sterile inoculation loop. Absorbance at 600 nm was measured in a spectrophotometer in order to determine bacterial optical density (OD_{600}). The volume of re-suspended cells to inoculate was calculated based on a starting OD_{600} of between 0.001 and 0.01. Cultures were grown overnight in 25 mL MSgg broth in a 250 mL conical flask in a water bath set to 37^{o}C with shaking at 130 rpm. Cultures at an OD_{600} ≥2 were harvested and normalized to 5 mL of the culture with the lowest OD_{600} (or alternatively normalization was carried out to the equivalent of 5 mL of culture at an OD_{600} of 2). The harvested culture was then pelleted by centrifugation at 3220 x g for 10 min and the full volume of supernatant was filter-sterilized to remove bacterial cells. The supernatant protease activity assay was set up by mixing supernatant 1:1 (vol:vol) with purified TapA protein and incubating the mixture at 37^{o}C for 24 h. This gave a final TapA protein concentration in the

assay of 5 µg/µL, and a total amount of 150 µg. Heat-inactivation of the supernatant was carried out by incubating the supernatant samples at 100^{o}C for 10 min prior to use.

2.4.10. Protease plate assays

Protease plate assays were carried as previously published (Verhamme et al., 2007). The production of exoproteases was assayed by the use of LB agar plates supplemented with 1.5% (w/v) dried milk powder. Strains of interest were each grown in 3 mL of LB broth at 37^{o}C to an OD_{600nm} of at least 1. The strains were then normalized to an OD_{600nm} of 0.01 by dilution in LB broth. 10 µL of the prepared cell culture was spotted onto the LB milk agar plate and grown for 18 h at 37^{o}C prior to being photographed.

2.4.11. Protein identification

Identification of proteins from biofilm lysates was carried out by first running samples by SDS-PAGE and then visualising protein by staining the gel with InstantBlue and then cutting out the band of interest with a sterile scalpel before sending for analysis. Proteins were identified by LC-MS-MS following tryptic digest and peptides were identified using the Mascot database. Mass spectrometry analysis was carried out by the FingerPrints Proteomics and Mass Spectrometry Facility at the University of Dundee.

2.4.12. Edman sequencing

Briefly, the sample was prepared for analysis by separating 30 µg of $TapA_{44-253}$ protein (after 84 days of storage at 4^{o}C) by SDS-PAGE on a 14% gel. The protein was then transferred to an immobilon-P PVDF membrane. The membrane was stained with InstantBlue to allow for the detection of a strongly staining band, corresponding to the stable TapA degradation product, which

was then cut out using a sterile scalpel and provided for analysis. Thanks to Dave Campbell and Bob Gourlay of the MRC protein Phosphorylation Unit in Dundee for Edman analysis carried out on a Shimadzu PPSQ-30 protein sequencer.

2.5. RNA methodology

The following procedure was carried out to confirm the expression of the *tapA* gene in strain NRS5787 (PY79 protease-free KO7) received from the *Bacillus* genetic stock centre (BGSC).

2.5.1. RNA extraction

Cultures were grown overnight in 25 mL MSgg broth in a 250 mL conical flask at 37^{o}C at 130 rpm from a starting OD_{600} of 0.01. Cells were harvested by centrifugation at an OD_{600} of between 1.71 and 1.9 and the amount of cells used for downstream RNA extraction normalized to the equivalent of 3 mL of cells at an OD_{600} equal to1. RNA extraction was carried out using the RiboPure™ RNA Purification Kit (Ambion) and the manufacturer's instructions were followed. Briefly, this method involves the use of beads to lyse the cells and a phenol-chloroform precipitation method to isolate RNA from other cellular material. To remove potential DNA carryover from the purification process DNase I treatment was employed using Ribopure DNaseI (Ambion). 4 U of DNaseI were incubated with 100 µL of RNA, from the previous step, alongside the corresponding DNase buffer and incubated at 37^{o}C for 30 min. DNase inactivation reagent was then used to stop the activity of DNaseI. The RNA quality was then assessed by measuring RNA concentration using the Nanodrop ND-1000 then approximately 150 ng of RNA was resolved on a 2%

(w/v) agarose gel by electrophoresis. High quality RNA was confirmed by the presence of 2 distinct bands corresponding to ribosomal RNA and by the lack of smearing indicative of a sample that has not significantly degraded.

2.5.2. Synthesis of complementary DNA (cDNA)

cDNA synthesis was carried out using the SuperScript™ III First-Strand Synthesis System (Invitrogen). Briefly, a mixture of 300 ng of RNA, 50 ng of hexamer primers and 1 mM mixed dNTPs were incubated at 65^{o}C for 5 min and then held on ice for at least 1 min. The cDNA synthesis mastermix was then added to each reaction to make a solution containing 5 mM $MgCl_2$, 10 uM DTT, 200 U Supercript III, 1 x RT buffer and 40 U of RNase out and made up to 20 µL with nuclease-free water. The reactions were then placed in a thermocycler on the following program: 10 min at 25^{o}C, 50 min at 50^{o}C and 5 min at 85^{o}C. Thereafter cDNA was stored at -20^{o}C.

2.5.3. Semi-quantitative reverse transcription (RT) PCR

To confirm successful removal of DNA from the purified RNA and the synthesis of cDNA from an RNA template a PCR was carried out using primers which anneal to the 16S *rRNA* gene (Primers DEN5 and DEN7 see **Table 9**) using g2 green mastermix PCR in a 20 µL PCR reaction, for the programme outlined in **Table 4**. On confirmation of cDNA synthesis semi-quantitative RT-PCR of the *tapA* gene was carried out using primers NSW 528 and NSW 529, using 100 ng of template cDNA. Then 8 µL or 12 µL, for the 16S and the *tapA* PCR amplification products were run on 1% (w/v) agarose gel electrophoresis and visualised using the Gel Doc XR system (Biorad).

Step	Temperature	Time
Initial denaturation	95^{o}C	2 min
Denaturation	95^{o}C	15 sec
Annealing	56^{o}C for 16S and 50^{o}C for *tapA*	30 sec
Extension	72^{o}C	1 min (30 sec per kb of DNA)
Final extension	72^{o}C	5 min
Hold	4-10^{o}C	

Table 4: The PCR conditions used in the thermocycler for semi-quantitative reverse transcription (RT) PCR. The Denaturation to Extension steps were cycled 35 times for 16S gene amplification and 35 times for *tapA* gene amplification.

2.6. Bioinformatics

Analysis of the *tapA* operon was carried out by using the online server ARNold to predict the presence of transcriptional terminators (Gautheret and Lambert, 2001; Hofacker et al., 1994; Lesnik et al., 2001; Macke et al., 2001). Comparative analysis of *tapA* operons was carried out using the Jalview 2 workbench (Waterhouse et al., 2009). Briefly, *tapA*, *sipW* and *tasA* ORFs (**Table 5**), originating from related *Bacillus* species, were analysed by pairwise sequence alignment with reference to the *B. subtilis* sequences to determine percentage sequence identity.

TapA protein homologs were identified using BlastP (Altschul et al., 1990; Altschul et al., 1997). Searching was carried out within the non-redundant protein sequence database for each of the organisms of interest using the amino acid sequence of the TapA protein originating from *B. subtilis* as the query. Protein sequences of TapA homologues were aligned using Clustal Omega using default settings (Sievers et al., 2011). Percentage identity between TapA orthologues was calculated with reference to *B. subtilis* TapA using the pairwise alignment function on the Jalview 2 workbench. For the prediction of signal peptides for all TapA variants then the SignalP 4.1 server

was used and set to the organism group: Gram-positive (Petersen et al., 2011). Secondary structure prediction was carried out using the JPred 4 server (Drozdetskiy et al., 2015). The prediction of TapA tertiary structure was carried out using the Phyre^2 (Kelley et al., 2015) and RaptorX (Källberg et al., 2012; Ma et al., 2012; Ma et al., 2013; Peng and Xu, 2010, 2011a, b) servers. The visualisation of the experimentally determined crystal structures of TasA and α-synuclein were carried out by loading the corresponding data file from the protein databank into the PyMol programme, as was the visualisation of the predicted tertiary structures of TapA (DeLano, 2016).

Genome description	**GenBank Genome accession number**
B. subtilis NCIB3610	CP020102.1
B. amyloliquefaciens FZB42	CP000560.1
*B. paralicheniformis**	MIZE01000002.1
B. pumilis SAFR-032	CP000813.4
B. cereus NC7401	AP007209

Table 5: GenBank accession numbers for the genomes in which the *tapA* homologs used in this study originate from. (*) The genetic material used to amplify the *B. paralicheniformis tapA* homolog was originally annotated as originating from *B. licheniformis*, however, on searching on the Blastn database with the sequence result of this amplicon it aligned more closely with the *B. paralicheniformis* sequences. The sequence was very similar to the *B. licheniformis* sequence expected but with some minor differences, note that this sequence may vary from the actual *tapA* sequence for this *B. paralicheniformis* strain as the primers were designed for *B. licheniformis* and likely introduced *B. licheniformis* specific *tapA* sequence elements (such as at the ribosome binding site and the start/ends of the open reading frame).

2.7. Buffers, solutions and reagents

Buffer/solution	Components
Generating competent 168 *B. subtilis* and transformation	
SpC	1 X T-Base 1 mM MgSO4 0.5% (w/v) Glucose
SpII	1 X T-Base 35 mM MgSO4 0.5% (w/v) Glucose 0.1% (w/v) Yeast extract 0.01% (w/v) Casamino Acids 40 µg/ml Tryptophan
Wash buffer	1X T-Base 1 mM Mg2SO4
Transformation buffer	1 X T-Base 2 mM EDTA
General solutions	
TAE Buffer	40 mM Tris 1.142% (w/v) Acetate 1 mM EDTA
DNA Loading Buffer (NEB)	2.5%(w/v) Ficoll®-400 11mM EDTA 3.3 mM Tris-HCl 0.017% (w/v) SDS 0.015% (w/v) bromophenol blue pH 8
B. subtilis biofilm lysis buffer: BugBuster protein extraction reagent (Novagen)	Contains detergent rLysozyme solution and Bezonase (nuclease)
Buffers for protein work	
Sample loading buffer for SDS-PAGE	60 mM Tris-HCl (pH 8.8) 4% (v/v) β-Mercaptoethanol 2% (w/v) SDS 10% (v/v) Glycerol 0.04% (w/v) Bromophenol Blue
NuPage Loading dye and reducing agent	Contains: glycerine lithium dodecyl sulphate coomassie G250 phenol red 50 mM dithiothreitol (DTT)
SDS Running Buffer	25 mM Tris 192 mM Glycine 0.1% (v/v) SDS
InstantBlue protein stain	Coomassie dye Phosphoric acid Solubilizing agents
Purification buffer	Tris 25 mM NaCl 250 mM (pH 7.6 with concentrated HCl)
Immunoblot (Western blot)	
Western transfer buffer	25 mM Tris 192 mM Glycine 0.2% (v/v) Tween-20 20% (v/v) Methanol
Tris Buffered Saline Tween (TBST)	20 mM Tris-HCl pH 8.0 150 mM NaCl 0.2% (v/v) Tween
ECL	Biorad Clarity Contains: Peroxide and Luminol enhancer

	reagents to be mixed 1:1
Miscellaneous	
Enzymatic lysis buffer (lysis of *B. subtilis* cells)	20 mM Tris-HCl at pH 8.0 2 mM sodium EDTA 1.2% (v/v) Triton X-100 20 mg/mL lysozyme

Table 6: This table lists the buffers, solutions and reagents used.

2.8. Strains, plasmids and primers used in this study

2.8.1. Strain table

Strain	Relevant genotype/description [a]	Source/construction [b]
168	*trpC2*	BGSC
NCIB3610	*prototroph*	BGSC
NRS3789	168 pMiniMAD Δ*tapA*	pNW685→168
NRS5048	168 pMiniMAD Δ*tasA*	pNW1448→168
NRS2012	NCIB3610 *sinIR::spec*	(Kearns et al., 2005)
NRS3936	NCIB3610 *ΔtapA*	Laura Hobley (This work)
NRS5267	NCIB3610 *ΔtasA*	(Erskine et al., 2018)
NRS5248	NCIB3610 *ΔtasA sinR*	SPP1 NRS5048→NCIB3610
NRS5487	168 pMiniMAD *ΔsipW*	pNW2021→168
NRS5488	NCIB3610 *ΔsipW*	SPP1 NRS5487→NCIB3610
NRS1144	*B. subtilis* RO-FF-1	(Roberts and Cohan, 1995)
NRS5142	*B. subtilis* ATCC 9799	(Duthie, 1944) BGSC:3A14
NRS5145	*B. subtilis* B-14393T	(Priest et al., 1988) BGSC:10A5
NRS1314	NCIB3610 *degU::pBL201 (cml)*	(Verhamme et al., 2007)
NRS5041	168 *amyE::P-spank-tapA*$_{B\_sub*}$*-lacI (spc)*	pNW1438→168
NRS5042	168 *amyE::P-spank-tapA*$_{B\_pum**}$*-lacI (spc)*	pNW1439→168
NRS5043	168 *amyE::P-spank-tapA*$_{B\_amy***}$*-lacI (spc)*	pNW1440→168
NRS5741	168 *amyE::P-spank-tapA*$_{B\_para****}$*-lacI (spc)*	pNW1800→168
NRS5742	168 *amyE::P-spank-tapA*$_{B\_sub\_1-193}$*-lacI (spc)*	pNW1801→168
NRS5045	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{B\_sub}$*-lacI (spc)*	SPP1 NRS5041→NRS3936 Generated by Rachel Gillespie
NRS5046	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{B\_pum}$*-lacI (spc)*	SPP1 NRS5042→NRS3936 Generated by Rachel Gillespie
NRS5047	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{B\_amy}$*-lacI (spc)*	SPP1 NRS5043→NRS3936 Generated by Rachel Gillespie
NRS5743	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{B\_para}$*-lacI (spc)*	SPP1 NRS5741→NRS3936
NRS5744	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1-193}$*-lacI (spc)*	SPP1 NRS5742→NRS3936
NRS5749	NCIB3610 *ΔtapA ΔtasA sinR*	SPP1 NRS3789→NRS5248
NRS5760	168 *amyE::P-spank-tapA-sipW-tasA-lacI (spc)*	pNW1840→168
NRS5763	NCIB3610 *ΔtapA ΔtasA sinR + amyE::P-spank-tapA-sipW-tasA-lacI (spc)*	SPP1 NRS5760→NRS5749
NRS5770	168 *amyE::P-spank-tapA*$_{Bs\_1-188aa}$*-lacI (spc)*	pNW1806→168
NRS5789	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{B\_sub\_1-188aa}$*-lacI (spc)*	SPP1 NRS5770→NRS3936
NRS5771	168 *amyE::P-spank-tapA*$_{Bs\_1-183aa}$*-lacI (spc)*	pNW1807→168
NRS5790	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{B\_sub\_1-183aa}$*-lacI (spc)*	SPP NRS15771→NRS3936
NRS5772	168 *amyE::P-spank-tapA*$_{Bs\_1-178aa}$*-lacI (spc)*	pNW1808→168
NRS5791	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1-178aa}$*-lacI (spc)*	SPP1 NRS5772→NRS3936
NRS5785	168 *amyE::P-spank-tapA*$_{Bs\_1-173aa}$*-lacI (spc)*	pNW1810→168
NRS5793	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1-173aa}$*-lacI (spc)*	SPP1 NRS5785→NRS3936
NRS5799	168 *amyE::P-spank-tapA*$_{Bs\_1-133aa}$*-lacI (spc)*	pNW1815→168
NRS5805	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1-133aa}$*-lacI (spc)*	SPP1 NRS5799→NRS3936
NRS5800	168 *amyE::P-spank-tapA*$_{Bs\_1-143aa}$*-lacI (spc)*	pNW1816→168
NRS5806	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1-143aa}$*-lacI (spc)*	SPP1 NRS5800→NRS3936
NRS5801	168 *amyE::P-spank-tapA*$_{Bs\_1-153aa}$*-lacI (spc)*	pNW1817→168
NRS5813	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1-153aa}$*-lacI (spc)*	SPP1 NRS5801→NRS3936

Strain	Relevant genotype/description [a]	Source/construction [b]
NRS5802	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}163aa}$*-lacI (spc)*	pNW1818→168
NRS5814	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}163aa}$*-lacI (spc)*	SPP1 NRS5802→NRS3936
NRS5819	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}123aa}$*-lacI* (spc)	pNW1819→168
NRS5987	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}123aa}$*-lacI (spc)*	SPP1 NRS5819→NRS3936
NRS5985	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}113aa}$*-lacI (spc)*	pNW1820→168
NRS5988	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}113aa}$*-lacI (spc)*	SPP1 NRS5985→NRS3936
NRS5986	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}103aa}$*-lacI (spc)*	pNW1821→168
NRS5989	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}103aa}$*-lacI (spc)*	SPP1 NRS5986→NRS3936
NRS5996	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}50aa}$*-lacI (spc)*	pNW1822→168
NRS6002	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}50aa}$*-lacI (spc)*	SPP1 NRS5996→NRS3936
NRS5997	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}71aa}$*-lacI (spc)*	pNW1823→168
NRS6003	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}71aa}$*-lacI (spc)*	SPP1 NRS5997→NRS3936
NRS5998	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}88aa}$*-lacI (spc)*	pNW1824→168
NRS6004	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}88aa}$*-lacI (spc)*	SPP1 NRS5998→NRS3936
NRS5999	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}95aa}$*-lacI (spc)*	pNW1825→168
NRS6005	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}95aa}$*-lacI (spc)*	SPP1 NRS5999→NRS3936
NRS6024	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}56aa}$*-lacI (spc)*	pNW1829→168
NRS6025	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}56aa}$*-lacI (spc)*	SPP1 NRS6024→NRS3936
NRS6036	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}57aa}$*-lacI (spc)*	pNW1830→168
NRS6041	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}57aa}$*-lacI (spc)*	SPP1 NRS6036→NRS3936
NRS6037	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}58aa}$*-lacI (spc)*	pNW1831→168
NRS6042	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}58aa}$*-lacI (spc)*	SPP1 NRS6037→NRS3936
NRS6038	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}59aa}$*-lacI (spc)*	pNW1832→168
NRS6043	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}59aa}$*-lacI (spc)*	SPP1 NRS6038→NRS3936
NRS6039	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}60aa}$*-lacI (spc)*	pNW1833→168
NRS 6044	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}60aa}$*-lacI (spc)*	SPP1 NRS6039→NRS3936
NRS6040	168 *amyE::P-spank-tapA*$_{Bs\_1\text{-}65aa}$*-lacI (spc)*	pNW1834→168
NRS6045	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{Bs\_1\text{-}65aa}$*-lacI (spc)*	SPP1 NRS6040→NRS3936
NRS5995	168 *amyE::P-spank-lacI (spc)*	pDR110→168
BGSC1A747	*B. subtilis* strain PY79	BGSC stocked as NRS5786
PY79-KO1	PY79 Δ*nprE* (KO1)	BGSC stocked as NRS5807
PY79-KO2	PY79 Δ*nprE Δ*aprE (KO2)	BGSC stocked as NRS5808
PY79-KO3	PY79 Δ*nprE Δ*aprE*Δ epr* (KO3)	BGSC stocked as NRS5809
PY79-KO4	PY79 Δ*nprE Δ*aprE Δepr Δmpr* (KO4)	BGSC stocked as NRS5810
PY79-KO5	PY79 Δ*nprE Δ*aprE Δepr Δmpr Δ*nprB* (KO5)	BGSC stocked as NRS5811
PY79-KO6	PY79 Δ*nprE Δ*aprE Δepr Δmpr Δ*nprB Δvpr* (KO6)	BGSC stocked as NRS5812
PY79-KO7	PY79 Δ*nprE Δ*aprE Δepr Δmpr Δ*nprB Δvpr Δbpr* (KO7)	BGSC stocked as NRS5787
NRS5788	168 *amyE::P-spank-tapA*$_{\Delta cons}$*-lacI (spc)*	pNW1814→168
NRS5794	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{\Delta cons}$*-lacI (spc)*	SPP1 NRS5788→NRS3936
NRS6058	168 *amyE::P-spank-tapA*$_{B\_sub\_A44P}$*-lacI (spc)*	pNW1838→168
NRS6059	NCIB3610 *ΔtapA + amyE::P-spank-tapA*$_{B\_sub\_A44P}$ *-lacI (spc)*	SPP1 NRS6058→NRS3936
NRS6379	168 *amyE::P-spank-tapA(rbs)-tasA(ss)*$_{1\text{-}27aa}$*-*	pNW1828→168

Strain	Relevant genotype/description [a]	Source/construction [b]
	tapA$_{44-57}$	
NRS6478	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA(rbs)-tasA(ss)*$_{1-27aa}$*-tapA*$_{44-57aa}$	NRS6020→NRS3936
NRS6489	168 *amyE::P-spank-tapA*$_{B\_pum1-40aa}$	pNW1875→168
NRS6508	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_pum1-40aa}$	SPP1 NRS6489→NRS3936
NRS6490	168 *amyE::P-spank-tapA*$_{B\_pum1-41aa}$	pNW1876→168
NRS6509	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_pum1-41aa}$	SPP1 NRS6490→NRS3936
NRS6491	168 *amyE::P-spank-tapA*$_{B\_pum1-42aa}$	pNW1877→168
NRS6510	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_pum1-42aa}$	SPP1 NRS6491→NRS3936
NRS6492	168 *amyE::P-spank-tapA*$_{B\_pum1-43aa}$	pNW1878→168
NRS6511	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_pum1-43aa}$	SPP1 NRS6492→NRS3936
NRS6481	168 *amyE::P-spank-tapA*$_{B\_amy1-57aa}$	pNW1871→168
NRS6504	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_amy1-57aa}$	SPP1 NRS6481→NRS3936
NRS6482	168 *amyE::P-spank-tapA*$_{B\_amy1-58aa}$	pNW1872→168
NRS6505	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_amy1-58aa}$	SPP1 NRS6482→NRS3936
NRS6483	168 *amyE::P-spank-tapA*$_{B\_amy1-59aa}$	pNW1873→168
NRS6506	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_amy1-59aa}$	SPP1 NRS6483→NRS3936
NRS6484	168 *amyE::P-spank-tapA*$_{B\_amy1-60aa}$	pNW1874→168
NRS6507	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_amy1-60aa}$	SPP1 NRS6484→NRS3936
NRS6497	168 *amyE::P-spank-tapA*$_{B\_para1-57aa}$	pNW1879→168
NRS6512	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_para1-57aa}$	SPP1 NRS6497→NRS3936
NRS6498	168 *amyE::P-spank-tapA*$_{B\_para1-58aa}$	pNW1880→168
NRS6513	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_para1-58aa}$	SPP1 NRS6498→NRS3936
NRS6499	168 *amyE::P-spank-tapA*$_{B\_para1-59aa}$	pNW1881→168
NRS6514	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_para1-59aa}$	SPP1 NRS6499→NRS3936
NRS6500	168 *amyE::P-spank-tapA*$_{B\_para1-60aa}$	pNW1882→168
NRS6515	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_para1-60aa}$	SPP1 NRS6500→NRS3936
NRS5990	168 *bpr::erm trpC2*	BKE15300 (Koo et al., 2017a)
NRS6010	168 *vpr::erm trpC2*	BKE38090 (Koo et al., 2017a)
NRS6011	168 *nprB::erm trpC2*	BKE11100 (Koo et al., 2017a)
NRS6012	168 *mpr::erm trpC2*	BKE02240 (Koo et al., 2017a)
NRS6013	168 *epr::erm trpC2*	BKE38400 (Koo et al., 2017a)
NRS6014	168 *aprE::erm trpC2*	BKE10300 (Koo et al., 2017a)
NRS6015	168 *nprE::erm trpC2*	BKE14700 (Koo et al., 2017a)
NRS6017	NCIB3610 *comI*$_{Q12L}$	BGSC 3A38
NRS6046	NCIB3610 *comI bpr::erm*	SPP1 NRS6046→NRS6017
NRS6047	NCIB3610 *comI Δbpr* (KO1)	pDR244→NRS6046
NRS6048	NICB3610 *comI Δbpr vpr::erm*	gDNA 6010→NRS6047
NRS6049	NCIB3610 *comI Δbpr Δvpr* (KO2)	pDR244→NRS6048
NRS6060	NCIB3610 *comI Δbpr Δvpr nprB::erm*	gDNA 6011→NRS6049
NRS6061	NCIB3610 *comI Δbpr Δvpr ΔnprB* (KO3)	pDR244→NRS6060
NRS6062	NCIB3610 *comI Δbpr Δvpr ΔnprB mpr::erm*	SPP1 NRS6012→NRS6061
NRS6063	NCIB3610 *comI Δbpr Δvpr ΔnprB Δmpr* (KO4)	pDR244→NRS6062
NRS6064	NCIB3610 *comI Δbpr Δvpr ΔnprB Δmpr epr::erm*	gDNA 6013→NRS6063
NRS6065	NCIB3610 *comI Δbpr Δvpr ΔnprB Δmpr Δepr* (KO5)	pDR244→NRS6065
NRS6340	NCIB3610 *comI Δbpr Δvpr ΔnprB Δmpr Δepr aprE::erm*	gDNA 6014→NRS6063
NRS6341	NCIB3610 *comI Δbpr Δvpr ΔnprB Δmpr Δepr ΔaprE* (KO6)	pDR244→NRS6340
NRS6361	NCIB3610 *comI Δbpr Δvpr ΔnprB Δmpr Δepr ΔaprE nprE::erm*	gDNA 6015→NRS6063
NRS6362	NCIB3610 *comI Δbpr Δvpr ΔnprB Δmpr Δepr ΔaprE ΔnprE* (KO7)	pDR244→NRS6361
NRS6366	168 *amyE::P-spank-tapA*$_{B\_sub\_1-57aa(T57A)}$*-lacI (spc)*	pNW1849→168
NRS6384	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_sub\_1-57aa(T57A)}$*-lacI (spc)*	SPP1 NRS6366→NRS3936
NRS6367	168 *amyE::P-spank-tapA*$_{B\_sub\_1-57aa(Q56A)}$*-lacI (spc)*	pNW1850→168
NRS6385	NCIB3610 *ΔtapA* + *amyE::P-spank-tapA*$_{B\_sub\_1-}$	SPP1 NRS6367→NRS3936

Strain	Relevant genotype/description [a]	Source/construction [b]
	$_{57aa(Q56A)}$-lacI (spc)	
NRS6368	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(L55I)}$-lacI (spc)*	pNW1851→168
NRS6386	NCIB 3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(L55I)}$-lacI (spc)*	SPP1 NRS6368→NRS3936
NRS6369	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(L55K)}$-lacI (spc)*	pNW1852→168
NRS6387	NCIB3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(L55K)}$-lacI (spc)*	SPP1 NRS6369→NRS3936
NRS6373	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(L55A)}$-lacI (spc)*	pNW1858→168
NRS6472	NCIB3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(L55A)}$-lacI (spc)*	SPP1 NRS6373→NRS3936
NRS6370	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(S54A)}$-lacI (spc)*	pNW1853→168
NRS6388	NCIB3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(S54A)}$-lacI (spc)*	SPP1 NRS6370→NRS3936
NRS6371	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(V53I)}$-lacI (spc)*	pNW1854→168
NRS6389	NCIB3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(V53I)}$-lacI (spc)*	SPP1 NRS6371→NRS3936
NRS6479	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(V53K)}$-lacI (spc)*	pNW1865→168
NRS6502	NCIB3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(V53K)}$-lacI (spc)*	SPP1 NRS6479→NRS3936
NRS6374	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(V53A)}$-lacI (spc)*	pNW1859→168
NRS6473	NCIB3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(V53A)}$-lacI (spc)*	SPP1 NRS6374→NRS3936
NRS6377	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(D52A)}$-lacI (spc)*	pNW1866→168
NRS6476	NCIB3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(D52A)}$-lacI (spc)*	SPP1 NRS6377→NRS3936
NRS6378	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(D52L)}$-lacI (spc)*	pNW1867→168
NRS6477	NCIB3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(D52L)}$-lacI (spc)*	SPP1 NRS6378→NRS3936
NRS6501	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(D52N)}$-lacI (spc)*	pNW1869→168
NRS6516	NCIB3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(D52N)}$-lacI (spc)*	SPP1 NRS6501→NRS3936
NRS6372	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(F51A)}$-lacI (spc)*	pNW1857→168
NRS6390	NCIB3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(F51A)}$-lacI (spc)*	SPP1 NRS6372→NRS3936
NRS6376	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(D47A)}$-lacI (spc)*	pNW1863→168
NRS6475	NCIB3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(D47A)}$-lacI (spc)*	SPP1 NRS6376→NRS3936
NRS6375	168 *amyE::P-spank-tapA$_{B_sub_1-57aa(F45A)}$-lacI (spc)*	pNW1862→168
NRS6474	NCIB3610 *ΔtapA + amyE::P-spank-tapA$_{B_sub_1-57aa(F45A)}$-lacI (spc)*	SPP1 NRS6375→NRS3936
NRS2705	168 *amyE::P-spank- sipW-tasA-lacI (spc)*	pNW1619→168
NRS6519	NCIB3610 *ΔtapA + amyE::P-spank- sipW- tasA-lacI (spc)*	SPP1 NRS6519→NRS3936
NRS6520	NCIB3610 *amyE::P-spank- tapA$_{B_sub_1-57aa(L55K)}$-lacI (spc)*	SPP1 NRS6369→NCIB3610
NRS6521	NCIB3610 *amyE::P-spank- tapA$_{B_sub_1-57aa(L55A)}$-lacI (spc)*	SPP1 NRS6373→NCIB3610
NRS6522	NCIB3610 *amyE::P-spank- tapA$_{B_sub_1-57aa(F45A)}$-lacI (spc)*	SPP1 NRS6375→NCIB3610
MC1061	*F' lacIQ lacZM15 Tn10 (tet)*	*E. coli* Genetic Stock Centre
BL21	F–*ompT hsdSB (rB–), (mB–)gal dcm* (DE3)	(Studier and Moffatt, 1986)
NRS2661	*E. coli* BL21 pGEX-6P-1+TEV-site+*tapA$_{B_sub}$*$_{44-253aa}$	pNW1600→*E.coli* BL21
NRS5034	*E. coli* BL21 pGEX-6P-1+TEV-site+*tapA$_{B_sub}$*$_{34-253aa}$	pNW1441→*E.coli* BL21
NRS5028	*E. coli* BL21 pGEX-6P-1+TEV site+*tasA$_{B\ sub28-261aa}$*	(Erskine et al., 2018)
NRS6016	*E. coli* CDR 1055 pDR244	BGSC ECE274

Table 7: Strain Table. a) Drug resistance cassettes are indicated as the following: (*spc)* spectinomycin resistance, (*erm*) erythromycin resistance and (*cml*) chloramphenicol resistance. b) BGSC is the *Bacillus* genetic stock centre. c) The direction of strain construction is indicated

with an arrow (→) from the genetic source either phage (SPP1), plasmid (pNW) or genomic DNA (gDNA) to the recipient strain. *Allele of *tapA* amplified from NCIB3610 (B_sub or Bs), **Allele of *tapA* amplified from *Bacillus pumilis* strain SAFR-032 (B_pum), ***Allele of *tapA* amplified from *Bacillus amyloliquefaciens* strain FZB42 (B_amy) and ****Allele of *tapA* amplified from *Bacillus paralicheniformis* (B_para). Note that the *B. paralicheniformis* isolate was originally annotated as *B. licheniformis* but re-assigned based on the similarity of the *tapA* gene sequence to that of *B. paralicheniformis*.

2.8.2. Plasmid table

Plasmid	Description	Source
pDR110	*B. subtilis* integration vector for IPTG-induced expression	(Britton et al., 2002)
pGEX-6P-1	Vector for overexpression of GST-fused proteins	GE Healthcare
pMiniMAD	Temp sensitive allelic replacement vector	(Patrick and Kearns, 2008)
pNW685	pMinimad Δ*tapA*	This work (Laura Hobley)
pNW1448	pMinimad Δ*tasA*	(Erskine et al., 2018)
pNW2021	pMinimad Δ*sipW*	This work (Elliot Erskine)
pNW1804	pDR110- *tapA-sipW-tasA*	This work (Elliot Erskine)
pNW1438	pDR110-*tapA* $_{B\ sub*}$	Rachel Gillespie
pNW1439	pDR110-*tapA* $_{B\ pum**}$	Rachel Gillespie
pNW1440	pDR110-*tapA* $_{B\ amy***}$	Rachel Gillespie
pNW1800	pDR110-*tapA* $_{B\ para****}$	This work
pNW1801	pDR110-*tapA* $_{B\ sub\ 1-193aa}$	This work
pNW1806	pDR110-*tapA* $_{B\ sub\ 1-188aa}$	This work
pNW1807	pDR110-*tapA* $_{B\ sub\ 1-183aa}$	This work
pNW1808	pDR110-*tapA* $_{B\ sub\ 1-178aa}$	This work
pNW1810	pDR110-*tapA* $_{B\ sub\ 1-173aa}$	This work
pNW1818	pDR110-*tapA* $_{B\ sub\ 1-163aa}$	This work
pNW1817	pDR110-*tapA* $_{B\ sub\ 1-153aa}$	This work
pNW1816	pDR110-*tapA* $_{B\ sub\ 1-143aa}$	This work
pNW1815	pDR110-*tapA* $_{B\ sub\ 1-133aa}$	This work
pNW1819	pDR110-*tapA* $_{B\ sub\ 1-123aa}$	This work
pNW1820	pDR110-*tapA* $_{B\ sub\ 1-113aa}$	This work
pNW1821	pDR110-*tapA* $_{B\ sub\ 1-103aa}$	This work
pNW1825	pDR110-*tapA* $_{B\ sub\ 1-95aa}$	This work
pNW1824	pDR110-*tapA* $_{B\ sub\ 1-88aa}$	This work
pNW1823	pDR110-*tapA* $_{B\ sub\ 1-71aa}$	This work
pNW1834	pDR110-*tapA* $_{B\ sub\ 1-65aa}$	This work
pNW1833	pDR110-*tapA* $_{B\ sub\ 1-60aa}$	This work
pNW1832	pDR110-*tapA* $_{B\ sub\ 1-59aa}$	This work
pNW1831	pDR110-*tapA* $_{B\ sub\ 1-58aa}$	This work
pNW1830	pDR110-*tapA* $_{B\ sub\ 1-57aa}$	This work
pNW1829	pDR110-*tapA* $_{B\ sub\ 1-56aa}$	This work
pNW1822	pDR110-*tapA* $_{B\ sub\ 1-50aa}$	This work
pNW1814	pDR110- *tapA* $_{B\ sub\ 111-SKAAAAAAKL-120}$	This work
pNW1864	pDR110-*tapA* $_{B\ sub\ RBS}$ *-tasA* $_{B\ sub\ 1-27aa}$*-tapA* $_{B\ sub\ 44-57aa}$	GenScript
pNW1838	pDR110-*tapA* $_{B\ sub\ A44P}$	This work
pNW1871	pDR110-*tapA* $_{B\ amy\ 1-57aa}$	This work
pNW1872	pDR110-*tapA* $_{B\ amy\ 1-58aa}$	This work
pNW1873	pDR110-*tapA* $_{B\ amy\ 1-59aa}$	This work
pNW1874	pDR110-*tapA* $_{B\ amy\ 1-60aa}$	This work
pNW1875	pDR110-*tapA* $_{B\ pum\ 1-40aa}$	This work
pNW1876	pDR110-*tapA* $_{B\ pum\ 1-41aa}$	This work
pNW1877	pDR110-*tapA* $_{B\ pum\ 1-42aa}$	This work
pNW1878	pDR110-*tapA* $_{B\ pum\ 1-43aa}$	This work
pNW1879	pDR110-*tapA* $_{B\ para\ 1-57aa}$	This work
pNW1880	pDR110-*tapA* $_{B\ para\ 1-58aa}$	This work
pNW1881	pDR110-*tapA* $_{B\ para\ 1-59aa}$	This work
pNW1882	pDR110-*tapA* $_{B\ para\ 1-60aa}$	This work
pNW1849	pDR110-*tapA* $_{B\ sub\ 1-57(T57A)}$	This work
pNW1850	pDR110-*tapA* $_{B\ sub\ 1-57(Q56A)}$	This work
pNW1851	pDR110-*tapA* $_{B\ sub\ 1-57(L55I)}$	This work
pNW1852	pDR110-*tapA* $_{B\ sub\ 1-57(L55K)}$	This work
pNW1858	pDR110-*tapA* $_{B\ sub\ 1-57(L55A)}$	This work
pNW1853	pDR110-*tapA* $_{B\ sub\ 1-57(S54A)}$	This work
pNW1854	pDR110-*tapA* $_{B\ sub\ 1-57(V53I)}$	This work

Plasmid	Description	Source
pNW1865	pDR110-*tapA*$_{B\ sub\ 1\text{-}57(V53K)}$	This work
pNW1859	pDR110-*tapA*$_{B\ sub\ 1\text{-}57(V53A)}$	This work
pNW1866	pDR110-*tapA*$_{B\ sub\ 1\text{-}57(D52A)}$	This work
pNW1867	pDR110-*tapA*$_{B\ sub\ 1\text{-}57(D52L)}$	This work
pNW1869	pDR110-*tapA*$_{B\ sub\ 1\text{-}57(D52N)}$	This work
pNW1857	pDR110-*tapA*$_{B\ sub\ 1\text{-}57(F51A)}$	This work
pNW1863	pDR110-*tapA*$_{B\ sub\ 1\text{-}57(D47A)}$	This work
pNW1862	pDR110-*tapA*$_{B\ sub\ 1\text{-}57(F45A)}$	This work
pNW1600	pGEX-6P-1-*tapA*$_{B\ sub44\text{-}253}$	This work
pNW1804	pDR110-*tapA-sipW-tasA*	This work
pNW1619	pDR110-*sipW-tasA-lacI (spc)*	(Erskine et al., 2018)
pNW1600	pGEX-6P-1+TEV-site+*tapA*$_{B\_sub44\text{-}253aa}$	This work (Laura D'Ignazio)
pNW1441	pGEX-6P-1+TEV-site+*tapA*$_{B\_sub34\text{-}253aa}$	This work Rachel Gillespie
pNW1437	pGEX-6P-1+TEV-site+*tasA*$_{B\ sub28\text{-}261aa}$	(Erskine et al., 2018)

Table 8: The plasmids used in this work. *Allele of *tapA* amplified from NCIB 3610 (B_sub), **Allele of *tapA* amplified from *Bacillus pumilis* strain SAFR-032 (B_pum), ***Allele of *tapA* amplified from *Bacillus amyloliquefaciens* strain FZB42 (B_amy) and ****Allele of *tapA* amplified from *Bacillus paralicheniformis* (B_para). Note that the *B. paralicheniformis* isolate was originally annotated as *B. licheniformis* but re-assigned based on the similarity of the *tapA* gene sequence to that of *B. paralicheniformis*.

2.8.3. Primer table

Primer	Sequence 5' – 3'[a]	Use	Plasmid[b]
NSW2181	GCTAGGATCCGAAAATTTATATTTTCAAGCTTTTCATGATATTGAAACATTTG	Forward primer from $TapA_{B\_sub44aa}$	pNW1600 For:2181 Rev:1895
NSW 1895	GCATCTCGAG**TTATTA**CTGATCAGCTTCATTGC	Reverse primer from $TapA_{B\_sub253aa}$	TapA44-253 into pGEX-6P-1 (Laura D'Ignazio)
NSW1896	ATGCGTCGACTTTTACAGGAGGTAAGATATGTTTCG	Forward primer $TapA_{B\_sub*}$	pNW1438 For:1896 Rev:1897
NSW1897	ATGCGCATGC**TTA**CTGATCAGCTTCATTGC	Forward primer $TapA_{B\_sub*\_full-length}$	
NSW2511	ATGCGCATGCTTAGCATTTTGCAAGCCTCATAGGC	Reverse primer $TapA_{B\_sub\_1-188aa}$	pNW1806 Primers For:1896 Rev:2511
NSW2512	ATGCGCATGCTTACATAGGCTCCGACCACTCA	Reverse primer $TapA_{B\_sub\_1-183aa}$	pNW1807 Primers For:1896 Rev:2512
NSW2513	ATGCGCATGCTTACTCAAATGTACTGCCGTTT	Reverse primer $TapA_{B\_sub\_1-178aa}$	pNW1808 Primers For:1896 Rev:2513
NSW2514	ATGCGCATGCTTAGTTTGCCGGGTAGCCTGCCGGT	Reverse primer $TapA_{B\_sub\_1-173aa}$	pNW1810 Primers For:1896 Rev:2514
NSW2521	ATGCGCATGCTTAGATCACGTTCCCATCCTTTAACG	Reverse primer $TapA_{B\_sub\_1-133aa}$	pNW1815 Primers For:1896 Rev:2521
NSW2520	ATGCGCATGCTTAGCCGATTTGATTGGAGAC	Reverse primer $TapA_{B\_sub\_1-143aa}$	pNW1816 Primers For:1896 Rev:2520
NSW2519	ATGCGCATGCTTATTTCTTGGTCTCAATTTTATAAAG	Reverse primer $TapA_{B\_sub\_1-153aa}$	pNW1817 Primers For:1896 Rev:2519
NSW2518	ATGCGCATGCTTATTTAAATGCATAAATGCCGG	Reverse primer $TapA_{B\_sub\_1-163aa}$	pNW1818 Primers For:1896 Rev:2518
NSW2522	ATGCGCATGCTTAGGCATTTTCAAGCTTATGAAGC	Reverse primer $TapA_{B\_sub\_1-123aa}$	pNW1819 Primers For:1896 Rev:2522
NSW2523	ATGCGCATGCTTACCACTTTGATTTCTTAAGTTTCTCACC	Reverse primer $TapA_{B\_sub\_1-113aa}$	pNW1820 Primers For:1896 Rev:2523

NSW2524	ATGCGCATGCTTAATTTTCGAGCACAGCAAATAAGGCG	Reverse primer $TapA_{B\_sub\_1-103aa}$	pNW1821 Primers For:1896 Rev:2524
NSW2531	ATGCGCATGCTTATGTTTCAATATCATGAAAAGC	Reverse primer $TapA_{B\_sub\_1-50aa}$	pNW1822 Primers For:1896 Rev:2531
NSW2532	ATGCGCATGCTTAATCATAATGGCAGTTTTTATC	Reverse primer $TapA_{B\_sub\_1-71aa}$	pNW1823 Primers For:1896 Rev:2532
NSW2533	ATGCGCATGCTTATTTCGTATCCGTTTGATCTG	Reverse primer $TapA_{B\_sub\_1-88aa}$	pNW1824 Primers For:1896 Rev:2533
NSW2534	ATGCGCATGCTTAGAAAGGTGAGCATACAGTGC	Reverse primer $TapA_{B\_sub\_1-95aa}$	pNW1825 Primers For:1896 Rev:2534
NSW2550	ATGCGCATGCTTATTGAAGTGAGACATCAAATGTTTC	Reverse primer $TapA_{B\_sub\_1-56aa}$	pNW1829 Primers For:1896 Rev:2550
NSW2565	ATGCGCATGCTTACGTTTGAAGTGAGACATC	Reverse primer $TapA_{B\_sub\_1-57aa}$	pNW1830 Primers For:1896 Rev:2565
NSW2566	ATGCGCATGCTTAACACGTTTGAAGTGAGAC	Reverse primer $TapA_{B\_sub\_1-58aa}$	pNW1831 Primers For:1896 Rev:2566
NSW2567	ATGCGCATGCTTATTTACACGTTTGAAGTGAG	Reverse primer $TapA_{B\_sub\_1-59aa}$	pNW1832 Primers For:1896 Rev:2567
NSW2568	ATGCGCATGCTTAGTCTTTACACGTTTGAAGTGAG	Reverse primer $TapA_{B\_sub\_1-60aa}$	pNW1833 Primers For:1896 Rev:2568
NSW2569	ATGCGCATGCTTAATCTGTATGCTGAAAGTC	Reverse primer $TapA_{B\_sub\_1-65aa}$	pNW1834 Primers For:1896 Rev:2569
NSW2001	ATGCGTCGACCAAGGAGTTGGAGAATGAATGAAACAGTCGC	Forward primer $TapA_{B\_pum**}$	pNW1827 Primers For:2001 Rev:2002
NSW2002	ATGCGCATGC**TTA**AGATACCTTTTCTGGACACTTTGC	Reverse primer $TapA_{B\_pum**\_Full-length}$	
NSW2535	ATGCGCATGCTTAAGGCTGACAATGCTGATCTG	Reverse primer $TapA_{B\_pum\_1-53aa}$	pNW1827 Primers For:2001 Rev:2535
NSW2866	ATGCGCATGC**TTA**TGTTTTCATGACAAAGTTTG	Reverse primer $TapA_{B\_pum\_1-40aa}$	pNW1875 Primers For:2001 Rev:2866

NSW2867	ATGCGCATGC**TTA**ACATGTTTTCATGACAAAG	Reverse primer $TapA_{B\_pum\_1-41aa}$	pNW1876 Primers For:2001 Rev:2867
NSW2868	ATGCGCATGC**TTA**TTCACATGTTTTCATGACA	Reverse primer $TapA_{B\_pum\_1-42aa}$	pNW1877 Primers For:2001 Rev:2868
NSW2869	ATGCGCATGC**TTA**ATTTTCACATGTTTTCATGAC	Reverse primer $TapA_{B\_pum\_1-43aa}$	pNW1878 Primers For:2001 Rev:2869
NSW2506	ATGCGTCGACCTCGCTAGGAGGGTAGTCATGTTCC	Forward primer $TapA_{B\_para****}$	pNW1800 Primers For:2506 Rev:2507
NSW2507	ATGCGCATGCTCATGACGCTTCCCCGCTTTC	Reverse primer $TapA_{B\_para****full-length}$	
NSW2991	ATGCGCATGC**TTA**CGCTTTCATTGTCATGC	Reverse primer $TapA_{B\_para1-57}$	pNW1879 Primers For:2506 Rev:2991
NSW2992	ATGCGCATGC**TTA**GGCCGCTTTCATTGTCATGC	Reverse primer $TapA_{B\_para1-58}$	pNW1880 Primers For:2506 Rev:2992
NSW2993	ATGCGCATGC**TTA**AGAGGCCGCTTTCATTGTCATG	Reverse primer $TapA_{B\_para1-59}$	pNW1881 Primers For:2506 Rev:2993
NSW2994	ATGCGCATGC**TTA**TGAAGAGGCCGCTTTCATTGTC	Reverse primer $TapA_{B\_para1-60}$	pNW1882 Primers For:2506 Rev:2994
NSW1898	gcatGTCGACTTTTACAGGGGGTAAGGCATGTTCCGATTGTTGC	Forward primer $TapA_{B\_amy***}$	pNW1440 Primers For:1898 Rev:1899
NSW1899	atgcGCATGC**TTA**CTGATCAGTTTCAGCGTTTTTTTCATGTTCTTCC	Reverse primer $TapA_{B\_amy***\ full-lenght}$	
NSW2862	ATGCGCATGC**TTA**CGTTTGAACCTGGAGACTGAAGC	Reverse primer $TapA_{B\_amy1-57}$	pNW1871 Primers For:1898 Rev:2862
NSW2863	ATGCGCATGC**TTA**GCACGTTTGAACCTGGAGACTG	Reverse primer $TapA_{B\_amy1-58}$	pNW1872 For:1898 Rev:2863
NSW2864	ATGCGCATGC**TTA**TCCGCACGTTTGAACCTGGAGAC	Reverse primer $TapA_{B\_amy1-59}$	pNW1873 For:1898 Rev:2864
NSW2865	ATGCGCATGC**TTA**ATCTCCGCACGTTTGAACCTGG	Reverse primer $TapA_{B\_amy1-60}$	pNW1874 For:1898 Rev:2865
NSW2576	CGATGATACAAGCGCTcctTTTCATGATATTGAA	$TapA_{B\_sub\_A44P}$	pNW1838 SDM
NSW2577	TTCAATATCATGAAAAggaGCGCTTGTATCATCG	$TapA_{B\_sub\_A44P}$	

NSW2841	ATGCGCATGC**TTA**AGCTTGAAGTGAGACATC	$TapA_{B\_sub\_1\text{-}57(T57A)}$	pNW1849 For:1896 Rev:2841
NSW2842	ATGCGCATGC**TTA**CGTTGCAAGTGAGACATC	$TapA_{B\_sub\_1\text{-}57(Q56A)}$	pNW1850 For:1896 Rev:2842
NSW2843	ATGCGCATGC**TTA**CGTTTGAATTGAGACATC	$TapA_{B\_sub\_1\text{-}57(L55I)}$	pNW1851 For:1896 Rev:2843
NSW2844	ATGCGCATGC**TTA**CGTTTGCTTTGAGACATC	$TapA_{B\_sub\_1\text{-}57(L55K)}$	pNW1852 For:1896 Rev:2844
NSW2845	ATGCGCATGC**TTA**CGTTTGAAGTGCGACATC	$TapA_{B\_sub\_1\text{-}57(S54A)}$	pNW1853 For:1896 Rev:2845
NSW2846	ATGCGCATGC**TTA**CGTTTGAAGTGAGATATC	$TapA_{B\_sub\_1\text{-}57(V53I)}$	pNW1854 For:1896 Rev:2846
NSW2858	GATATTGAAACATTTGATAAGTCACTTCAAACGTG	$TapA_{B\_sub\_1\text{-}253(V53K)}$	pNW1855 For:2858 Rev: 2859 KOD (QC)
NSW2859	CACGTTTGAAGTGACTTATCAAATGTTTCAATATC	$TapA_{B\_sub\_1\text{-}253(V53K)}$	
NSW2849	ATGCGCATGC**TTA**CGTTTGTGCTGAGACATC	$TapA_{B\_sub\_1\text{-}57(L55A)}$	pNW1858 For:1896 Rev:2846
NSW2565	ATGCGCATGC**TTA**CGTTTGAAGTGAGACATC	$TapA_{B\_sub\_1\text{-}57(F51A)}$	pNW1857 For:1896 Rev:2565
NSW2850	ATGCGCATGC**TTA**CGTTTGAAGTGATGCATC	$TapA_{B\_sub\_1\text{-}57(V53A)}$	pNW1859 For:1896 Rev:2850
NSW2854	GATATTGAAACATTTGCTGTCTCACTTCAAACG	$TapA_{B\_sub\_1\text{-}253(D52A)}$	pNW1856 For:2854 Rev:2855 KOD (QC)
NSW2855	CGTTTGAAGTGAGACAGCAAATGTTTCAATATC	$TapA_{B\_sub\_1\text{-}253(D52A)}$	
NSW2584	GCGCTGCTTTTCATGCTATTGAAACATTTGATGTC	$TapA_{B\_sub\_1\text{-}253(D47A)}$	pNW1861 For:2584 Rev:2585 KOD (QC)
NSW2585	GACATCAAATGTTTCAATAGCATGAAAAGCAGCGC	$TapA_{B\_sub\_1\text{-}253(D47A)}$	
NSW2856	GATATTGAAACATTTAATGTCTCACTTCAAACG	$TapA_{B\_sub\_1\text{-}253(D52N)}$	pNW1868 For:2584 Rev:2585 KOD (QC)
NSW2857	CGTTTGAAGTGAGACTAAAAATGTTTCAATATC	$TapA_{B\_sub\_1\text{-}253(D52N)}$	
NSW2829	GATACAAGCGCTGCTGCTCATGATATTGAAACA	$TapA_{B\_sub\_1\text{-}253(F45A)}$	pNW1845 For:2829 Rev:2830 KOD (QC)
NSW2830	TGTTTCAATATCATGAGCAGCAGCGCTTGTATC	$TapA_{B\_sub\_1\text{-}253(F45A)}$	
NSW872	AGGTGTGGCATAATGTGTGTAATTGTGAGC	pDR110 seq For	Sequencing primers for pDR110
NSW873	TGAACAATCACGAAACAATAATTGGTACGTACG	pDR110 seq rev	
NSW528	CGATTGTTTCACAATCAG	*tapA* qPCR for	Primers for qRT-PCR of *tapA*
NSW529	CCACTTCCACTTTGATTT	*tapA* qPCR rev	
NSW2750	ACTCCTACGGGAGGCAGC	16S rRNA gene forward	16S amplification for phylogeny assignment
NSW2751	TCACGACACGAGCTGACGAC	16S rRNA gene reverse	
NSW1308	GCATGGATCCCTCTTCCCATTTGGACATGTG	*tapA* upstream for	Primers for the tapA

NSW1332	GGTAAGATATGTTTCGATTGGTCGACATGC	*tapA* upstream rev	deletion construct in the pMAD vector pNW685
NSW1333	GCATGTCGACCAGAAGGAAAGCGGGGAAGAG	*tapA* downstream for	
NSW1334	GCATGAATTCATATCGAAACCTGTTGCCAGG	*tapA* downstream rev	

Table 9: Primer table. a) Restriction sites are underlined, stop codons are in bold. b) QC = QuickChange (using KOD polymerase) and SDM = Site-Directed Mutagenesis. *Allele of *tapA* amplified from NCIB 3610 (B_sub), **Allele of *tapA* amplified from *Bacillus pumilis* strain SAFR-032 (B_pum), ***Allele of *tapA* amplified from *Bacillus amyloliquefaciens* strain FZB42 (B_amy) and ****Allele of *tapA* amplified from *Bacillus paralicheniformis*

3. Results

3.1. A biochemical approach to understand TapA function

3.1.1. Background

In a 1999 study (Stöver and Driks, 1999a), research was carried out investigating the *Bacillus subtilis* gene *yqxM,* which would later be renamed *tapA* (for TasA assembly/anchoring protein A) (Romero et al., 2011). The designation TapA will be used hereafter. At that time, the function of the *tapA* gene was unknown. TapA was not detected in cells grown in rich media or in a number of other conditions which were tested. It was reported that TapA could be detected in the supernatant, but not the cell extracts, of *B. subtilis* PY79 cells grown in LB supplemented with NaCl (sodium chloride) (Stöver and Driks, 1999a). In this case TapA was found to migrate as a protein of approximately 38 kDa. Next, by using PY79 cells where the *sipW* gene was deleted, it was shown that the secretion of TapA was dependent on the SipW signal peptidase. It was also found that when *tapA* expression was controlled by an IPTG-inducible P_{spac} promoter then the protein was secreted and found at a size of 30 kDa. It was noted that full-length TapA produced recombinantly in *E. coli* was found at a molecular weight of 38 kDa. Finally, when both SipW and TapA were produced together in the same *E. coli* cells there was evidence that SipW processes TapA, in the form of a shift to a lower molecular weight of TapA, as detected by immunoblot.

Investigating TapA in the conext of biofilm formation it was discovered that TapA localizes to both the cell and matrix fractions of pellicle biofilms grown in MSgg medium at the 24 h time point (Romero et al., 2011). In this case TapA was detected by immunoblot as 2 bands with a larger band of 28 kDa and a smaller band of 24 kDa. After 48 hs, only the smaller band remained detectable

in the matrix fraction only. The authors concluded that the 28 kDa band was equivalent to TapA without a signal sequence (which they predict as being from amino acids 1-33) and that the band of 24 kDa was indicative of further processing (Romero et al., 2011). Next a number of immuno-histochemical techniques, in combination with microscopy led to the conclusion that TapA is found as localised puncta associated with the cell surface and is found in TasA fibres anchored to the cell envelope. Additionally, it was found by immunoblot that TasA levels are low in biofilms formed by the Δ*tapA* mutant (Romero et al., 2011).

The aim of this section of work was to first confirm some of the findings in relation to research into TapA function. We then took a fresh look starting with the generation of an αTapA antibody to allow for the detection of TapA *in vivo*. With the aid of an αTapA antibody, it was found that TapA is processed to a low molecular weight form *in vivo* and that this is a conserved feature of several *B. subtilis* isolates. Finally, it was demonstrated that processing of TapA is dependent upon the action of secreted proteases found in the NCIB3610 biofilm environment.

3.1.2. The *tapA* gene is essential for biofilm formation

It has been reported that biofilm formation by *B. subtilis* is dependent on both *tapA* and *tasA* (Branda et al., 2006; Hamon et al., 2004). Here first, it was determined whether deletion of the Δ*tapA* (NRS3936) and Δ*tasA* (NRS5267) genes, separately, impacted biofilm formation as reported. Wild-type 3610, Δ*tapA* (NRS3936), Δ*tasA* (NRS5267) and ∆*tapA* + *pIPTG-tapA* (NRS5045) strains were grown on MSgg media for 48 h before image capture. Our results were in agreement with previous findings as deletion of either gene resulted in flat, featureless biofilms in contrast to the rugose, complex morphology of wild-type 3610 (**Figure 11**A). To determine if the phenotype of the Δ*tapA* mutant was due to the deletion of *tapA* and not due to off-target mutations, we examined the ability of ectopically expressed *tapA* to complement this biofilm defect. To test this, the expression of *tapA* was induced at the *amyE* locus, under the control of the IPTG-inducible P_{spank} promoter, using the strain Δ*tapA* + *pIPTG-tapA* (NRS5045) (**Figure 11**A). The induction of *tapA* gene expression was shown to restore wild-type biofilm architecture (**Figure 11**A). These results demonstrated that both *tasA* and *tapA* encode products which are essential for the ability of *B. subtilis* to form biofilms as previously reported. We also noted that the morphology of the *tapA* and *tasA* deletion strains were reproducibly different.

3.1.3. TapA is needed for the stability of TasA levels in the biofilm

As shown above both Δ*tapA* and Δ*tasA* mutants have biofilm defects, additionally, published work has shown that TasA levels are found to be reduced in a ∆*tapA* mutant (Romero et al., 2011). For this reason we wanted to examine TasA levels in the ∆*tapA* mutant generated in our laboratory.

The aim in the following experiment was to determine whether *tapA* deletion impacts the broad protein profile in the biofilm, therefore, strains 3610 and Δ*tapA* (NRS3936) were grown as biofilms on MSgg agar for 48 h. Protein was extracted from the biofilms and the biofilm lysates were analysed by SDS-PAGE. A strongly staining band of approximately 30 kDa was present in wild-type biofilms but was absent in the Δ*tapA* sample (**Figure 11**B, Arrow denoted *). It was predicted that this band could either be TasA or TapA as both proteins are predicted to have similar molecular weights of 28.31 kDa and 29.08 kDa respectively. To identify the protein, this band was isolated and confirmed to contain TasA by mass spectrometric analysis (**Figure 11**C). This result indicated that the loss of this band in the *tapA* mutant is due to reduced levels of TasA**.** For further validation, the biofilm lysates from a selection of mutants were analysed by αTasA immunoblot. The stained SDS-PAGE gel shown in **Figure 11**D serves as a loading control for this experiment. The result of the αTasA immunoblot demonstrates that in 3610 biofilms, TasA is found as a band of >25 kDa with a laddering of smaller forms found at a lower molecular weight below, which may be degraded forms of TasA. There was no band detected in the Δ*tasA* biofilm as expected (**Figure 11**E). SinR is known to repress *tasA* transcription and when the *sinR* gene is disrupted this increases *tasA* transcription (Chu et al., 2006). Correspondingly, there are increased levels of TasA in the *sinIR::spec* (NRS2012) mutant and TasA is not detected in the Δ*tasA sinR* lysates. Importantly, there are reduced levels of TasA observed in Δ*tapA* biofilms. Furthermore, TasA levels were shown to be recovered on induction of *tapA* expression when the strain Δ*tapA* + *pIPTG-tapA* was grown in the presence of IPTG (**Figure 11**E). The presence of degraded forms of TasA is also shown to be dependent on the level of *tapA* induction. The result of these

experiments are in agreement with previously published work that *tapA* is needed for the stability or production of TasA levels in the biofilm. As a result of

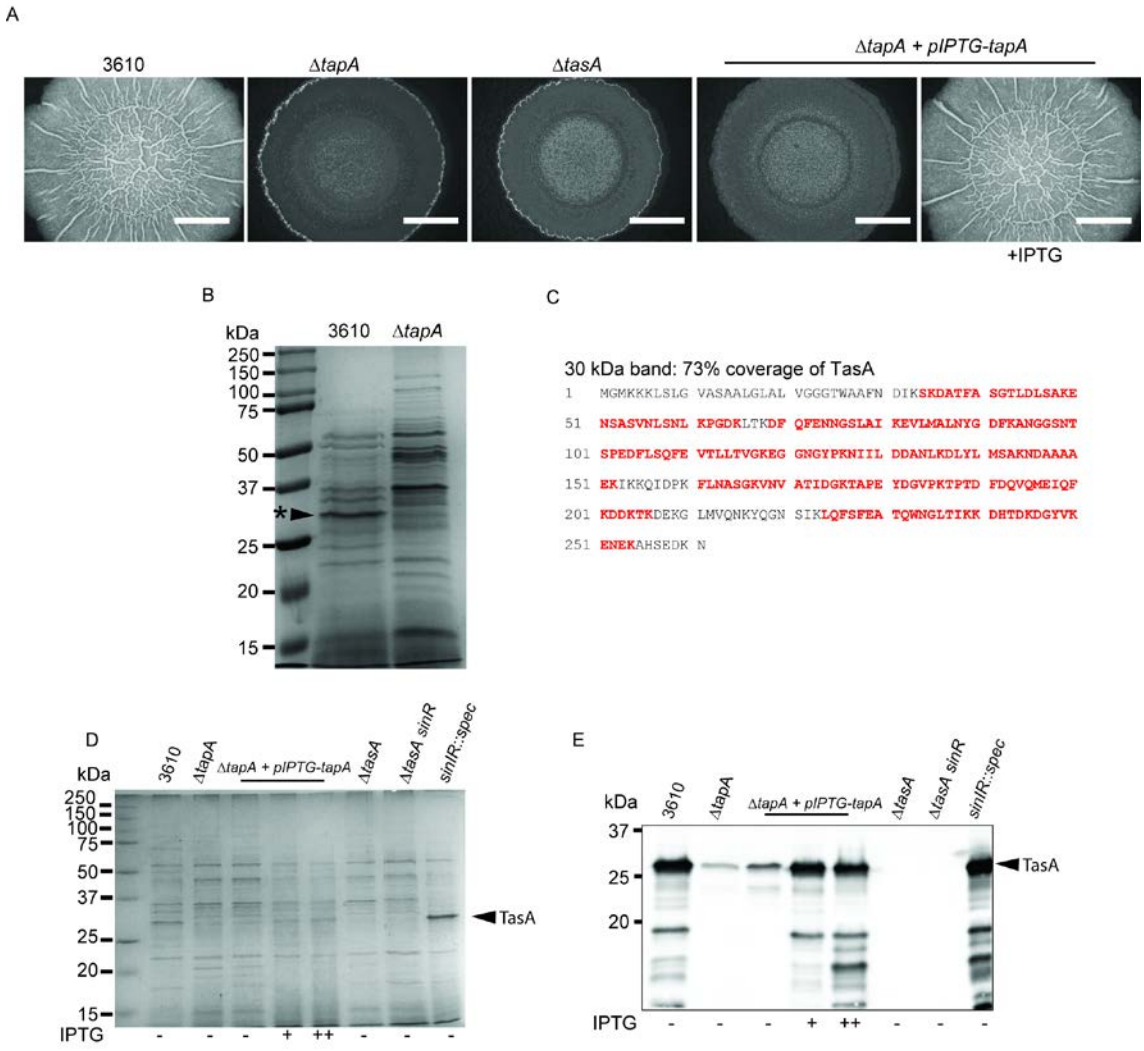


Figure 11: TapA is needed for TasA stability or production *in vivo.* (A) Biofilm phenotypes of: 3610, Δ*tapA* (NRS3936), Δ*tasA* (NRS5267) and Δ*tapA* + *pIPTG-tapA* (NRS5045) strains. Δ*tapA* + *pIPTG-tapA* (NRS5045) was grown in the presence and absence of IPTG. Biofilms were grown at 30^{o}C for 48 h. n= 3+. **(B)** Biofilm lysates of 3610 and Δ*tapA* were analysed by 12% SDS-PAGE, the gel is stained with InstantBlue, the maximal amount of protein was loaded into the well for each sample to increase the amount of material submitted for mass spectrometry analysis. The arrow denoted * indicates a 30 kDa band identified by mass spectrometry as TasA (C). **(C)** The results of mass spectrometry based protein identification. The protein sequence of TasA is shown and peptides detected which match the sequences are highlighted in red and bold text. n=1. **(D)** Biofilm lysates of 3610, Δ*tapA* (NRS3936), Δ*tapA* + *pIPTG-tapA* (NRS5045), Δ*tasA* (NRS5267), Δ*tasA sinR (*NRS5248) and *sinIR::spec* (NRS2012) were analysed by SDS-PAGE. An image of the 12% gel stained with InstantBlue serves as a loading control for the αTasA immunoblot shown in E. **(E)** αTasA immunoblot of selected biofilm lystaes. n=1. - = 0 μM IPTG + = 25 μM IPTG and ++ = 50 μM IPTG. Scale bars = 0.5 cm.

these findings, the Δ*tapA* mutant may be thought of as somewhat akin to a Δ*tapA* Δ*tasA* double mutant, in that it lacks the *tapA* gene and has reduced TasA levels.

3.1.4. Increased induction of *tasA* in a Δ*tapA* mutant does not restore biofilm formation

It was next tested if increasing TasA levels in a Δ*tapA* mutant could re-instate biofilm formation. To test this, a second copy of *tasA* was introduced into the *amyE* locus of the Δ*tapA* mutant to generate the strain Δ*tapA* + *pIPTG-sipW-tasA* (NRS6519). In an attempt to boost *tasA* levels, increasing concentrations of IPTG (25 µM to 1 mM) were used to induce *tasA* expression from the IPTG inducible promoter. The resulting biofilms did not display wild-type morphology but were distinguishable from both the Δ*tapA* and Δ*tasA* mutants (**Figure 12**A). To determine whether increased induction of *tasA* expression manifested as higher levels of TasA protein the biofilm lysates were analysed by αTasA immunoblot. The stained SDS-PAGE gel shown in **Figure 12**B serves as a loading control for this experiment. In the 3610 sample, TasA was found with an apparent molecular weight of ~30 kDa and there were degraded forms of the protein found at a lower molecular weight below (**Figure 12**C). By comparison, there was a reduced level of TasA in the Δ*tapA* mutant as was shown previously (**Figure 12**C). The lysates of the Δ*tapA* + *pIPTG-sipW-tasA* (NRS6519) strain either grown in the absence or presence of 25 µM of IPTG had comparable levels of TasA to the Δ*tapA* mutant (**Figure 12**C). Increasing the levels of IPTG to 50 µM, 100 µM and 1 mM did not recover the levels of TasA to that observed in wild-type biofilms (**Figure 12**C). Additionally, the

presence of only one lower molecular weight form of TasA was observed below the 30 kDa TasA band, in the Δ*tapA* and Δ*tapA* + *pIPTG-sipW-tasA* samples (**Figure 12**C). This lower band appeared to be found at a higher intensity as IPTG increased. However, this result confirmed that TasA levels could not be re-instated to wild-type levels by increased induction of *tasA* expression in a Δ*tapA* mutant.

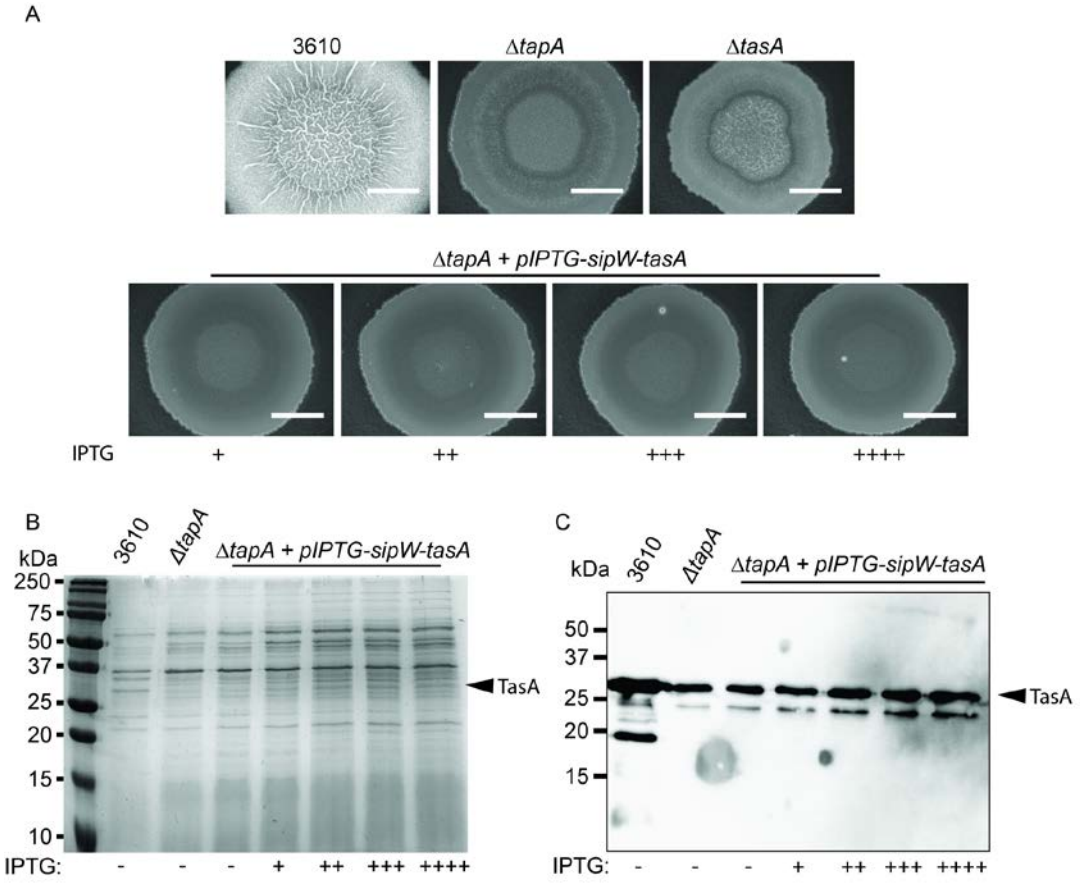


Figure 12: The impact of increased *tasA* induction, in the absence of *tapA*, on biofilm formation. (A) Biofilm phenotypes of 3610, Δ*tapA* and Δ*tapA* + *pIPTG-sipW-tasA* (NRS6519) grown at 30°C for 48 h. Δ*tapA* + *pIPTG-sipW-tasA* (NRS6519) was grown in the presence and absence of IPTG for induction of *tasA*. Biofilms were grown in the presence of varying IPTG concentrations. n=2. **(B)** Biofilms from the experiment shown in (A) were lysed and run on 14% SDS-PAGE which served as a loading control for the αTasA immunoblot shown in **(C)**. For (C) n=1. + = 25 µM IPTG, ++ = 50 µM IPTG, +++ = 100 µM IPTG and ++++ = 1 mM IPTG. Scale bars = 0.5 cm.

To summarise, we have reproduced published findings that a Δ*tapA* mutant displays a biofilm defect and that TasA levels are reduced in a Δ*tapA* mutant. Attempts to re-instate TasA levels by increased induction of *tasA* were unsuccessful.

3.1.5. TapA is found as a low molecular weight form in the biofilm

We have gained some insight to the function of *tapA* by deleting the gene and working with the Δ*tapA* mutant. However, to observe TapA *in vivo* it was determined that the generation of an αTapA antibody was required. The full-length TapA protein is calculated to have a molecular weight (MW) of 29.08 kDa. However, TapA is thought to be a secreted protein and consistent with this the SignalP server predicts a signal peptide of 43 amino acids (Petersen et al., 2011). Cleavage with a signal peptidase would yield a mature protein consisting of amino acids 44-253 giving a molecular weight of 24.18 kDa. For generation of an αTapA antibody against the mature secreted protein, the recombinant protein $TapA_{34-253}$ was produced in *E. coli* BL21 (DE3) (NRS5034) cells. Note that $TapA_{34-253}$ is the predicted mature form of the protein as previously published (Romero et al., 2014). Strain NRS5034 harbours the plasmid pGEX-6P-1-$tapA_{B\_sub34-253}$ (pNW1441) which encodes a GST (glutathione S-transferase) tag separated by a TEV (Tobacco Etch Virus) protease cleavage site from $TapA_{34-253}$. The GST-tag was used for the purification of $TapA_{34-253}$ before being removed by cleavage with TEV-protease to make tag-less $TapA_{34-253}$. The protein was then used to elicit antibody production in a rabbit host. The antibody was initially validated by immunoblot with biofilm lysates and was then subsequently affinity purified using the antigen $TapA_{44-253}$ as bait.

To test the specificity of the resulting antibody the following strains were used: 3610, Δ*tapA* (NRS3936), Δ*tapA* + *pIPTG-tapA* (NRS5045), Δ*tasA* (NRS5267), Δ*tasA sinR* (NRS5248) and *sinIR::spec* (NRS2012). The strains were grown under biofilm conditions for 48 h and whole biofilm protein extracts isolated. The stained SDS-PAGE which served as a loading control for this experiment is shown in **Figure 13**A. Three bands of interest were detected in 3610 lysates with the αTapA antibody. These were estimated to have an apparent molecular weight of 30 kDa (Arrow 1), 18 kDa (Arrow 2) and 16 kDa (Arrow 3) (**Figure 13**B). All 3 bands were absent in Δ*tapA* biofilm lysates. The detection of all 3 bands was shown to be IPTG-dependent in Δ*tapA* + *pIPTG-tapA* biofilm lysates (**Figure 13**B). Only the 16 kDa band was present in the ∆*tasA* and ∆*tasA sinR* biofilms. Therefore, it was concluded that the 16 kDa band is the only one that can be confidently assigned as TapA. All 3 bands were present in the *sinIR::spec* mutant with the 18 kDa band detected at a stronger intensity with comparison to wild-type (**Figure 13**B). In *sinIR::spec* cells, repression of the *tapA* operon is expected to be perturbed resulting in higher expression. Taken together, it can be concluded that the presence of the 18 kDa band is shown to be dependent on the presence of both the *tasA* and *tapA* genes and could possibly be TasA due to its absence in ∆*tasA* biofilm lysates (**Figure 13**B). The 30 kDa band detected by the αTapA antibody is anticipated to be a non-specific band corresponding to TasA as it is absent in the Δ*tasA* mutant. It is thought to be equivalent to the band, identified previously as TasA, which stains strongly with InstantBlue when looking at 3610 biofilm lysates by SDS-PAGE (**Figure 11**B). Given the unexpected nature of the findings, this result was then validated using an αTapA antibody that was previously published (Stöver and Driks, 1999a).

Comparable observations were made when αTapA serum was used (kindly provided by Adam Driks) (Stöver and Driks, 1999a). The same set of samples as shown in **Figure 13**B were analysed using this αTapA serum. 3 bands (Arrows 1, 2 and 3) were observed arranged in a similar banding pattern as detected with our own αTapA antibody (**Figure 13**C), thus validating the αTapA antibody we generated as part of this work. Additionally, there were a number of non-specific bands detected when the αTapA serum was used (**Figure 13**C). For example, an additional band of 15 kDa could be observed in 3610 biofilm lysates but this was found to be absent in a Δ*tasA* mutant and may represent TasA or a TasA-dependent form of TapA.

Following on from this, mass spectrometry was used in an attempt to identify the composition of the 16 kDa TapA form. To obtain high levels of 16 kDa TapA, *tapA* expression was induced with 100 μM IPTG in the strain Δ*tapA* + *pIPTG-tapA* (**Figure 13**D). The region of the SDS-PAGE gel, corresponding to the 16 kDa region where TapA is detected by immunoblot, was submitted for analysis (denoted arrow *). The quality of this result was low, likely due to the presence of other proteins in the sample but 3 peptides did match to the core of the TapA sequence (**Figure 13**E).

To conclude, an αTapA antibody was generated and shown to react non-specifically with TasA. However, a low molecular weight 16 kDa band is detected when TasA is absent from the sample and can, therefore, be assigned as TapA. This result indicated that TapA is processed to a low molecular weight form within the context of biofilm formation.

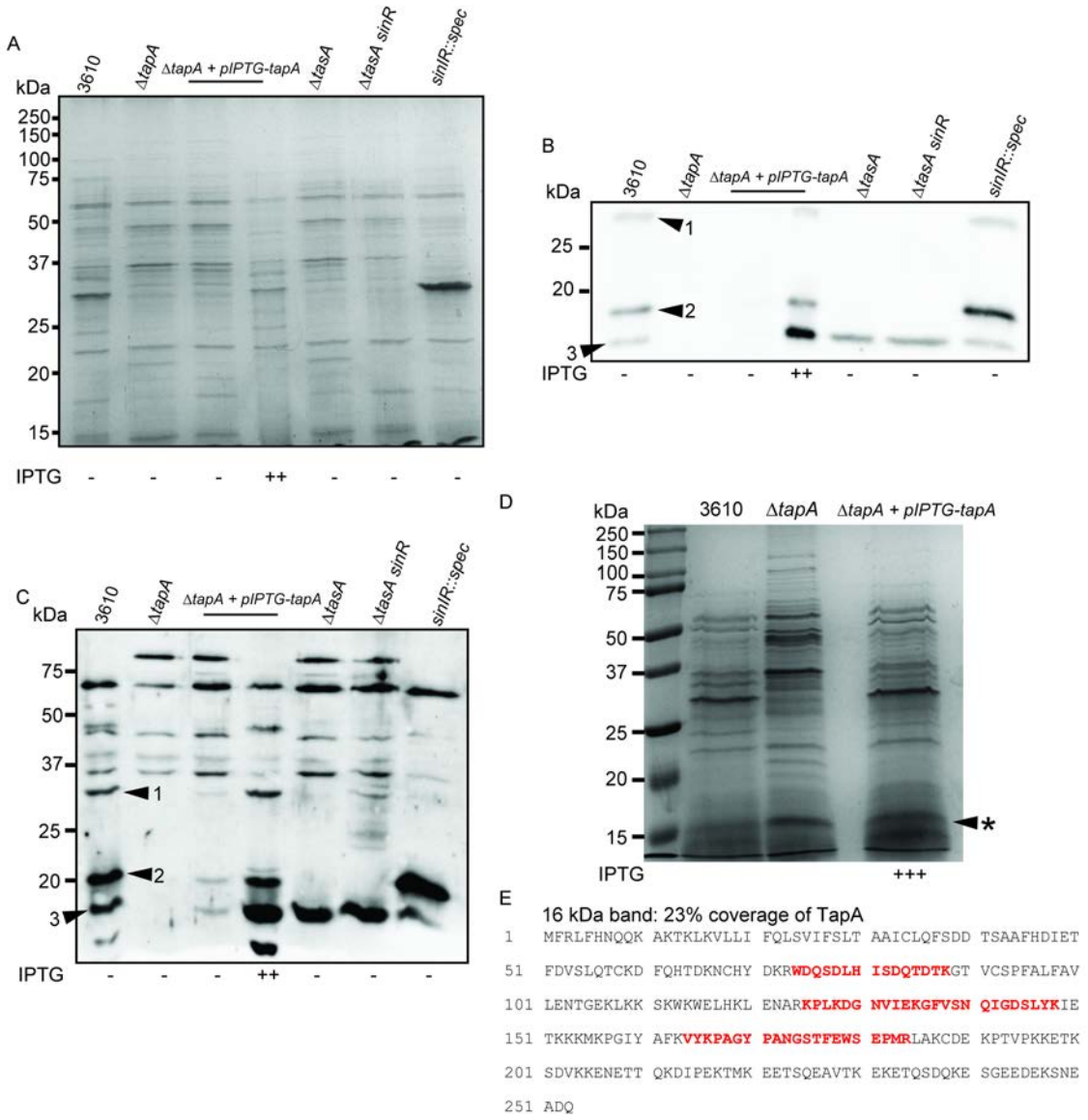


Figure 13: TapA is found as a low molecular weight form in biofilm lysates. (A) Biofilm lysates (~46 µg) were analysed on a 12% SDS-PAGE gel and stained with InstantBlue. Strains: 3610, Δ*tapA* (NRS3936), Δ*tapA* + *pIPTG-tapA* (NRS5045), Δ*tasA* (NRS5267), Δ*tasA sinR (*NRS5248) and *sinIR::spec* (NRS2012) are shown. The Δ*tapA* + *pIPTG-tapA* (NRS5045) strain was grown in the presence and absence of IPTG as indicated. The stained gel serves as the loading control for the immunoblots shown in (B) and (C). **(B)** αTapA immunoblot using the TapA antibody generated as part of the current study. n=2. **(C)** αTapA immunoblot using serum kindly gifted by Adam Driks. n=1. For the immunoblots shown in (B) and (C): arrow 1 corresponds to TasA (30 kDa band), arrow 2 indicates a TasA/TapA-dependent band (~18 kDa) and arrow 3 is TapA (16 kDa). **(D)** Biofilm lysates of 3610, Δ*tapA* and Δ*tapA* + *pIPTG-tapA* analysed by 12% SDS-PAGE, the gel is stained with InstantBlue, the maximal amount of protein was loaded into the well for each sample to increase the amount of material submitted for downstream analysis, Arrow denoted * indicates a 16 kDa band submitted for protein identification by mass spectrometry. Note that this is an extended version of the figure shown in **Figure 12**B. **(E)** Mass spectrometry result for the 16 kDa band shown in **(D)**. n=1. The protein

sequence of TapA is shown and peptides detected which match the sequences are highlighted in red and bold text. There was a limited number of peptides detected and TapA was not a leading hit for this band in the mass spectrometry results. - = 0 µM IPTG, ++ = 50 µM IPTG and +++ = 100 µM IPTG.

3.1.6. TapA processing is a conserved feature of *B. subtilis* isolates

TapA was found to be processed to a low molecular weight in *B. subtilis* 3610, however, it was not known whether this trait is specific to the 3610 strain. To test this, it was investigated whether TapA processing is a conserved feature across *B. subtilis* isolates. Three isolates were obtained from the *Bacillus* Genetic Stock Centre (BGSC). Strain RO-FF-1 was isolated from the Mojave Dessert (Roberts and Cohan, 1995), strain ATCC 9799 was first reported in 1944 as being obtained from a National Repository of Cultures (Duthie, 1944), and B-14393T has been published as an *B. amyloliquefaciens* strain (Priest et al., 1988). As a first step, these strains were identified phylogenetically using 16S rRNA sequencing and confirmed to be *B. subtilis* species (data not shown). This included the strain B-14393T which was previously assigned as *B. amyloliquefaciens*. To determine whether these *B. subtilis* isolates formed biofilms the strains were grown on MSgg media for 48 h before biofilm images were captured. Each of the isolates developed mature biofilms with complex architecture (**Figure 14**A). The biofilms exhibited by ATCC 9799 resembled the appearance of those formed by 3610, however, RO-FF-1 and B-14393T biofilms adopted different morphologes. Following on from this it was investigated whether TapA undergoes processing in these isolates. To do this biofilms were lysed and protein extracted for analysis by αTapA immunoblot. TapA processing was confirmed by the observation of a similar banding pattern in the 3 isolates when compared to the 3610 strain. This included the presence

of the low molecular weight TapA band (Arrow deonoted *) which is found in the Δ*tasA* strain and absent in Δ*tapA* (**Figure 14**B). These results demonstrated that TapA processing in the biofilm is a feature shared across the several isolates of *B. subtilis*.

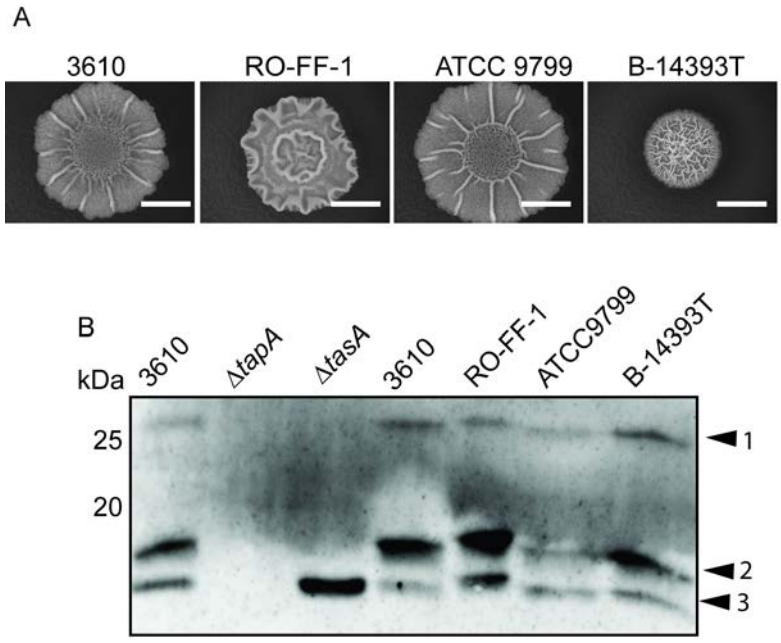


Figure 14: TapA processing is conserved in Bacillus subtilis species. (A) Biofilm images of *Bacillus subtilis* isolates 3610, RO-FF-1 (NRS1144), ATCC 9799 (NRS5142) and B-14393T (NRS5145) showing complex colony morphology at 48 h. images representative of at least 3 independent replicates. **(B)** αTapA immunoblot of biofilm lysates of the biofilms shown in (A). n=3. Biofilm lysate samples (~48 µg) were separated on 12% SDS-PAGE. The samples of 3610, Δ*tapA* (NRS3936) and Δ*tasA* (NRS5267) on the left hand-side of the blot originate from a separate experiment and are included for comparison. Arrow 1 corresponds to TasA (30 kDa band), arrow 2 indicates a TasA/TapA-dependent band (~18 kDa) and arrow 3 is TapA (16 kDa). Scale bar = 0.5 cm.

3.1.7. Extracellular proteases display proteolytic activity against TapA *in vitro*

As TapA is predicted to be a secreted protein, it was hypothesised that TapA processing in the biofilm would require the action of secreted extracellular proteases (referred to as exoproteases hereafter). As a starting point, it was tested whether exoproteases produced by *B. subtilis* and secreted into the external environment could cleave recombinant TapA *in vitro*. First, recombinant

$TapA_{44-253}$ was purified from *E. coli* BL21 (DE3) (NRS2661) cells and the protein sample was prepared under reducing and denaturing conditions. The $TapA_{44-253}$ protein was found to exhibit an apparent molecular weight of 25 kDa (**Figure 15**A). Higher molecular weight bands were observed and these may be oligomeric forms of TapA (**Figure 15**A). Next, to determine if recombinant TapA protein can be processed by exoproteases the following experiment was carried out. Briefly, wild-type 3610 cells were grown in MSgg broth to stationary phase. Cell-free supernatant was obtained by centrifugation of the culture and filtering of the supernatant. The supernatant was then incubated with 150 µg of the protein for 24 h at 37^{o}C. The $TapA_{44-253}$ protein, which served as the input for this assay, was found at an apparent molecular weight of ~25 kDa, as expected (**Figure 15**B). However, $TapA_{44-253}$ which had been incubated with 3610 supernatant was found as a degradation product of between 15 and 20 kDa (**Figure 15**B). Heat-inactivation of the supernatant at 100^{o}C for 10 min abolished proteolytic activity (3610 +H) (**Figure 15**B). It was noted that higher molecular weight forms of TapA were observed in the input sample and in the protein incubated with heat-treated supernatant (3610 +H). These results indicated that the degradation process observed is enzymatic in nature. The band corresponding to the protein core of approximately 17.5 kDa, left by proteolysis with cell-free supernatant, was isolated and analysed by mass spectrometry (**Figure 15**C). It was found that this corresponded to the core of the protein from amino acids 72-184. The resulting protein based on this mass spectrometry analysis has a molecular weight of only 12.98 kDa but it may be that there are smaller peptides not detected by the mass spectrometry analysis. This could explain the discrepancy in size between the mass spectrometry result and that which was observed by SDS-PAGE analyses. The result of this

experiment confirmed that 3610 supernatant had proteolytic activity against recombinant TapA protein.

3.1.8. $TapA_{44-253}$ is unstable *in vitro*

Consistent with TapA being subjected to proteolysis to a stable core, it was observed on multiple occasions that $TapA_{44-253}$ was found to be unstable when stored in purification buffer (pH 7.6), at $4^{o}C$. To test this systematically, a controlled experiment was set up to observe the degradation process over time. It was found that the degradation of $TapA_{44-253}$ progressed until a stable core was left. The stable core had an apparent molecular weight of ~17.5 kDa as judged by migration on SDS-PAGE (**Figure 15**D). As before the band was submitted for mass spectrometry analysis and found to comprise amino acids 72-230, which is calculated to correspond to a protein of 18.22 kDa (**Figure 15**E). Further analysis of the stable core by N-terminal sequencing, preliminarily indicated that the N-terminus of this protein started at amino acid 57 to give a protein of amino acids 57-230 which is predicted to be 20.05 kDa (**Figure 15**F). However, it should be noted that attempts to reproduce the result given by N-terminal sequencing have not yet been successful, this is likely because the sample submitted is a product of a degradation process, therefore, it might not be of the purity required to obtain reliable results. Interestingly, the stable core detected here is similar to that obtained from degradation of recombinant TapA protein subjected to proteolysis (**Figure 15**C).

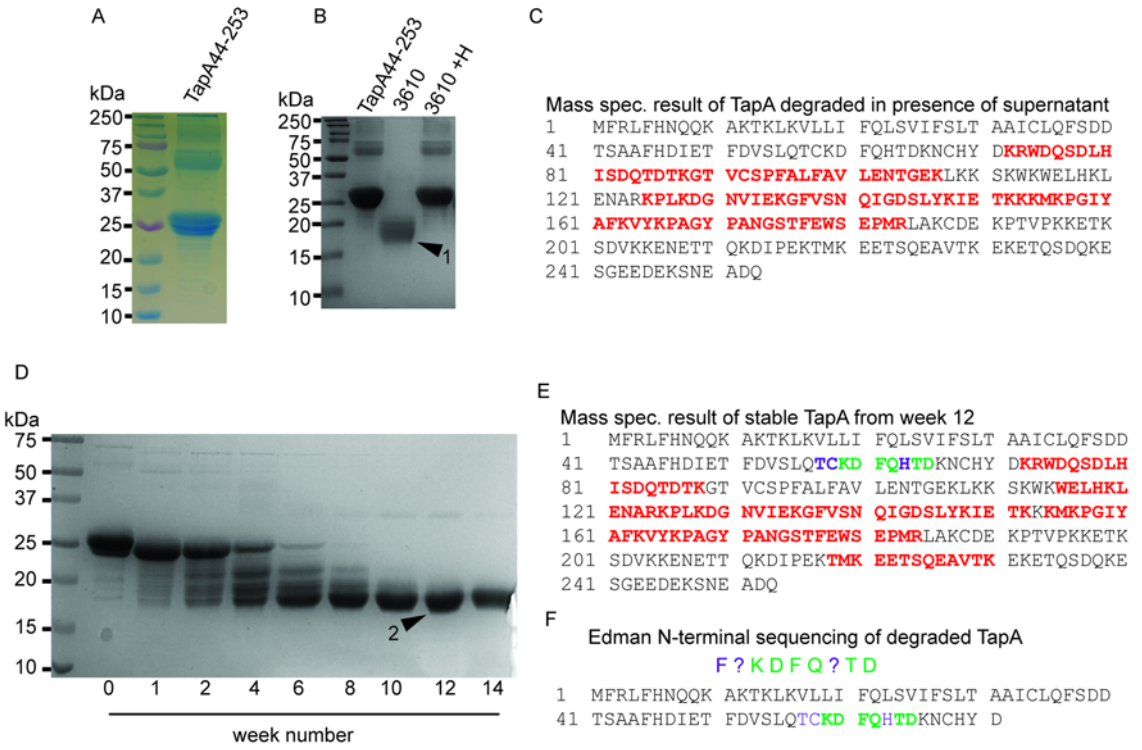


Figure 15: TapA is sensitive to degradation by extracellular enzymes. (A) $TapA_{44-253}$ was purified and eluted in purification buffer, lyophilized and stored at 4°C and then re-suspended in water. 30 µg of protein was prepared in reducing conditions and run on 14% SDS-PAGE. **(B)** $TapA_{44-253}$ protein (150 µg) was mixed 1:1 (vol:vol) with cell-free supernatant and incubated for 24 h at 37°C. Arrow 1 indicates the degraded protein submitted for analysis shown in (C). **(C)** Protein identification result for TapA subjected to proteolysis with cell-free supernatants as shown in B. n=1. **(D)** TapA protein stored at 4°C in purification buffer was sampled at intervals of 7 days. The protein was boiled and subject to reducing conditions by incubation with β-Me and prepared samples were then stored at -20°C. 28 µg of TapA protein was run on 14% SDS-PAGE. Arrow denoted 2 indicates the TapA stable core from week 12 which was submitted for protein identification. **(E)** Protein identification result for the TapA degradation product observed at week 12 shown in (C). n=1. **(F)** The result of Edman sequencing after 12 week old protein, the same sample analysed in (C), underwent N-terminal sequence analysis. n=1. Amino acids highlighted in green were consistent between the N-terminal sequencing result and the TapA sequence, amino acids in purple did not match. The sequence of TapA is shown with matching peptides identified by mass spectrometry highlighted in red and bold text.

3.1.9. The supernatant of PY79 has proteolytic activity against recombinant TapA

Having shown that recombinant TapA is subject to degradation by the proteolytic activity of 3610 supernatant *in vitro*, it was predicted that

exoproteases are responsible for TapA processing in biofilms. The aim of the following experiment was to determine whether extracellular proteases are responsible for the processing of TapA to a low molecular weight *in vivo*. Testing this in *B. subtilis* 3610 biofilms would involve extensive mutagenesis to delete all of the genes encoding exoproteases. For this reason we first utilized available tools in the form of a Δ7 exoprotease-free *B. subtilis* PY79 strain, which is freely available to the community via the BGSC. The PY79 exoprotease-free strain is denoted PY79-KO7 (**Table 10**) (Pohl et al., 2013). First, it was determined whether PY79 would serve as a viable model in place of 3610 by testing the proteolytic activity of PY79 supernatant against recombinant TapA. Briefly, cell-free culture supernatants were obtained from PY79 (NRS5786) and PY70-KO7 (NRS5787), which had been grown in MSgg media to stationary phase, and aliquots then incubated with 150 µg of $TapA_{44-253}$ at 37^{o}C for 24 h. When TapA was co-incubated with PY79 supernatant the protein was found to migrate on SDS-PAGE with a lower apparent molecular weight when compared with the $TapA_{44-253}$ that was used as the input in the assay (**Figure 16**A). To test whether this activity was exoprotease-dependent, TapA was co-incubated with PY79-KO7 supernatant. The product of this incubation was run on SDS-PAGE. In this case, TapA migrated to a similar extent as the input but migrated slightly further (**Figure 16**A). This indicated that TapA was subjected to a reduced level of proteolysis in the presence of PY79-KO7 supernatant. The proteolytic activity of both PY79 and PY79-KO7 supernatants was abolished with heat treatment at 100^{o}C for 10 min (PY79 +H and PY79-KO7 +H) (**Figure 16**A). To summarise, PY79 supernatant displayed proteolytic activity against recombinant TapA protein as has been shown above for the

supernatant of 3610. This proteolytic activity was perturbed in a PY79-KO7 mutant lacking genes encoding exoproteases.

3.1.10. The general exoprotease activity of PY79 is abolished in the strains PY79-KO2 to PY79-KO7

To determine which of the 7 exoproteases were important for TapA proteolytic activity the intermediate strains made as part of the project which generated PY79-KO7 were obtained from BGSC (**Table 10**). First, the exoprotease activity of PY79-KO7 and all of the intermediates strains PY79-KO1 to PY79-KO6 was tested. This was done by growth of the strains on LB agar, supplemented with 1.5% (w/v) milk for 18 h at 37^{o}C. Exoprotease activity can be observed in this assay by the presence of a “halo” around the growing colony which results from degradation of milk proteins present in the media. The strain 3610 *degU* was included as a negative control, this strain carries a disruption in the *degU* gene which regulates the expression of degradative enzymes (Cairns et al., 2015). 3610 *degU* did not display proteolytic activity in this assay. It was found that wild-type PY79 and PY79-KO1 displayed protease activity, as judged by the presence of a halo around the colony (**Figure 16**B). However, strains PY79-KO2 to PY79-KO7 did not display any observable proteolysis of milk proteins indicating reduced exoprotease activity (**Figure 16**B). This result demonstrated that the non-specific proteolytic activity, apparent by this assay, of PY79 can be perturbed in the absence of just two of the exoproteases, encoded by the *nprE* and *aprE* genes, which are deleted in the strain PY79-KO2. However, as multiple mutagenesis steps were carried out in order to generate strain PY79-KO7, further validation of this strain was required.

Strain	Genotype	strain number
PY79	Parental strain	NRS5786
PY79-KO1	*ΔnprE*	NRS5807
PY79-KO2	*ΔnprE ΔaprE*	NRS5808
PY79-KO3	*ΔnprE ΔaprE Δepr*	NRS5809
PY79-KO4	*ΔnprE ΔaprE Δepr Δmpr*	NRS5810
PY79-KO5	*ΔnprE ΔaprE Δepr Δmpr ΔnprB*	NRS5811
PY79-KO6	*ΔnprE ΔaprE Δepr Δmpr ΔnprB Δvpr*	NRS5812
PY79-KO7	*ΔnprE ΔaprE Δepr Δmpr ΔnprB Δvpr Δbpr*	NRS5787

Table 10: A list of the PY79 exoprotease mutant strains used in this study.

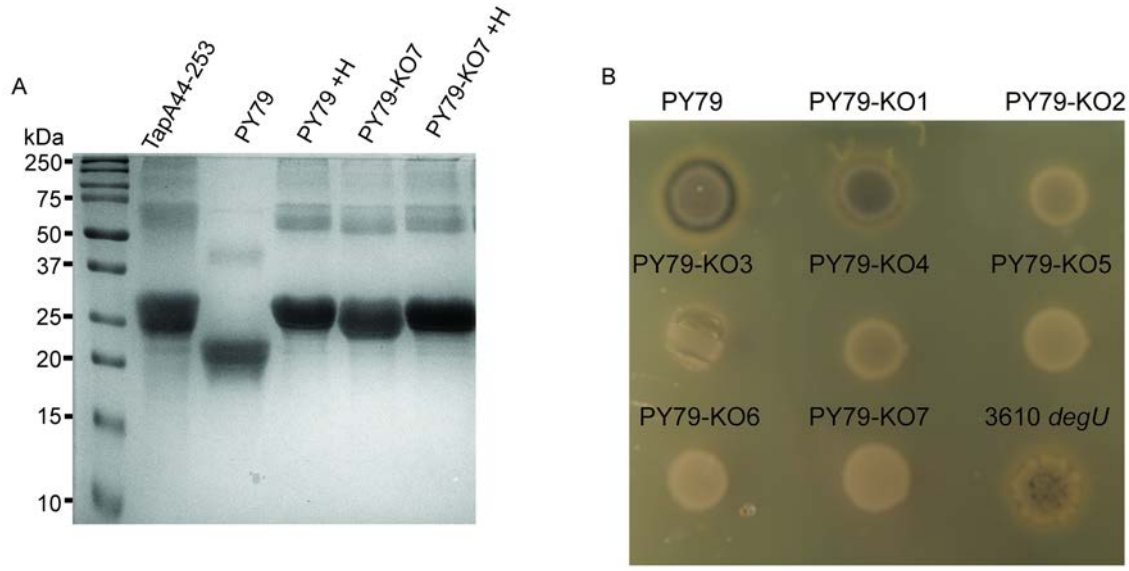


Figure 16: TapA processing by PY79 supernatant is exoprotease-dependent. (A) $TapA_{44-253}$ protein (150 µg) was mixed 1:1 (vol:vol) with cell-free supernatant and incubated for 24 h at 37°C. 30 µg of protein was run on 14% SDS-PAGE and stained with InstantBlue. The supernatant was incubated at 100°C for 10 min for heat-inactivation (+H). n=1. **(B)** PY79 (NRS5786) and intermediate protease mutants (**Table 10**) were grown on LB agar containing 1.5% milk. 3610 *degU* is included as an exoprotease-negative control. n=3.

3.1.11. The *tapA* gene is expressed in PY79-KO7

The PY79-derived strains used in this study have undergone multiple mutagenesis steps and PY79 does not form complex colony biofilms. Therefore, it was investigated as to whether *tapA* was transcribed in both the parental strain PY79 (NRS5786) and in the mutant strain PY79-KO7 (NRS5787). To test this, the strains 3610, 3610 Δ*tapA* (NRS3936), PY79 (NRS5786) and PY79-

KO7 (NRS5787) were grown in MSgg broth to stationary phase (**Figure 17**A). Total RNA was extracted from cell lysates and RNA quality was assesed by agarose gel electrophoresis (**Figure 17**B). The isolation of high quality RNA was confirmed by the observation of distinct bands corresponding to rRNA and by the absence of smearing which when present is indicative of degradation in each of the samples. Isolated RNA was then used as the template in the synthesis of cDNA with a reverse transcriptase enzyme. Genomic DNA isolated from PY79-KO7 (PY79-KO7 gDNA) served as the positive control in the subsequent PCR of *16S rRNA* and *tapA* (**Figure 17**C). The no template control was included to ensure the result of the PCR was not impacted by contaminating DNA sequences. It was confirmed by reverse-transcription PCR (RT-PCR) of the 16S rRNA gene that cDNA was successfully generated in each of the samples (**Figure 17**C). The 16S amplification was only observed in the samples which had been processed in the presence of the reverse transcriptase (RT) enzyme indicating that the RNA isolated was free of DNA contamination. To determine if the *tapA* transcript was present in the samples PCR amplification was carried out with *tapA* specific primers (**Figure 17**C). It was found that the 3610 sample contained *tapA* transcript which was absent in the Δ*tapA* sample, as expected. Both PY79 and PY79-KO7 were positive for the *tapA* transcript (**Figure 17**C). This confirms that the *tapA* gene is expressed in PY79 and PY79-KO7 grown under conditions known to induce *tapA* expression in 3610. This confirms the validity of using the PY79 and exoprotease mutants, which are derived from this strain, as an initial model for the investigation of TapA processing *in vivo*.

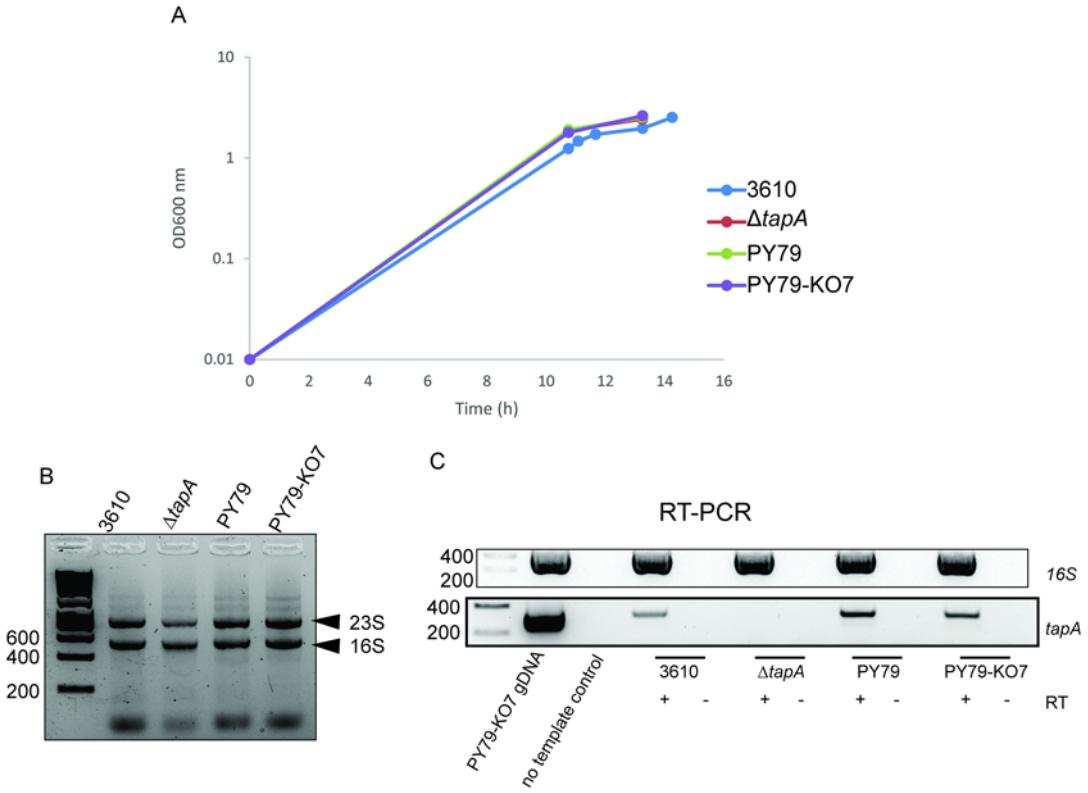


Figure 17: The *tapA* transcript is present in PY79 and PY79-KO7 cells grown to stationary phase. (A) Growth curve of 3610, Δ*tapA* (NRS3936), PY79 (NRS5786) and PY79-KO7 (NRS5787) strains grown in MSgg broth to stationary phase, as judged when growth levelled off at an OD_{600nm} measurement of approximately 2.5. The x-axis measure of OD_{600nm} is shown as a logarithmic scale, the y-axis shows time in hours (h). **(B)** RNA isolated from stationary phase cultures. >100 ng of purified RNA from each sample was run on a 2% agarose gel to assess RNA quality. **(C)** RT-PCR of (Top) 16S rRNA to confirm the presence of cDNA in RNA samples treated with reverse transcriptase (RT) and (Bottom) RT-PCR of *tapA* demonstrating the presence of *tapA* transcript in PY79-KO7 cells. n=1.

3.1.12. TapA processing is perturbed in exoprotease mutants derived from PY79

As demonstrated above, the *tapA* gene is expressed in the PY79 strain background, it was next investigated if the TapA protein could be detected *in vivo* and whether it undergoes processing in the absence of the exoprotease encoding genes. To test this, protein was extracted from cells grown under the same conditions as before and cell lysates were analysed by αTapA immunoblot. As was shown previously, 3 bands were detected by the αTapA antibody in the 3610 lysate and the same characteristic banding pattern was

observed in the PY79 lysate (**Figure 18**A). As expected these bands were absent in the Δ*tapA* lysate (**Figure 18**A). This result confirmed that PY79 cells produces TapA under the conditions tested. In PY79-KO7, the highest molecular weight TasA band of ~30 kDa was present but the other 2 bands were absent. Interestingly, an additional band, found below the TasA band, with an apparent molecular weight of ~28 kDa (Arrow *) was detected in the protease-free cells (**Figure 18**A). This band is similar in size to the calculated full-length TapA MW of 29.08 kDa (**Figure 18**A). This may represent an unprocessed form of TapA. These findings, in conjunction with the absence of the low molecular weight form of TapA and the TasA/TapA-dependent band, indicated that TapA processing is perturbed in PY79-KO7.

In order to determine whether cleavage of TapA was dependent on the action of any specific protease the intermediate exoprotease-mutant PY79 strains (**Table 10**), the TapA banding pattern was analysed by αTapA immunoblot. Strains were grown as before in MSgg media to stationary phase before cell lysates were separated by SDS-PAGE. It was found that TapA processing occurred similar to wild-type, with the detection of the low molecular weight 16 kDa form of TapA, in strains PY79-KO1 to PY79-KO5 (**Figure 18**B, arrow 3). The 16 kDa band was, however, absent in PY79-KO6 and the protease-free strain PY79-KO7 (**Figure 18**B). Therefore, the loss of the low molecular weight form of TapA coincided with the deletion of *vpr*. In contrast, the TasA/TapA-dependent band found at an apparent molecular weight of 18 kDa was present in all strains, with the exception of PY79-KO7 (**Figure 18**B, arrow 2). In PY79-KO7 the 28 kDa band, which is thought to be unprocessed TapA, was observed as before (**Figure 18**B, arrow 1). To summarise, PY79-KO6 and PY79-KO7 strains displayed the most noticeable differences in TapA processing. This indicated

that either there is redundancy in exoprotease activity against TapA or that the final 2 exoprotease genes deleted have a specific role in processing TapA.

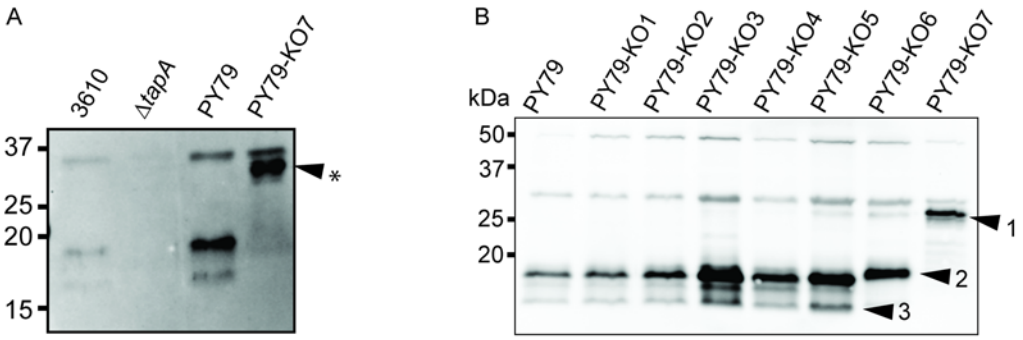


Figure 18: TapA processing in PY79 exoprotease mutants as analysed by αTapA immunoblot. (A) αTapA immunoblot of 3610, Δ*tapA* (NRS5045), PY79 (NRS5786) and PY79-KO7 (NRS5787) grown in MSgg broth. Arrow * = unprocessed TapA. n=1. **(B)** αTapA immunoblot of PY79 (NRS5786) and PY79 exoprotease mutant cell lysates: PY79-KO1 (NRS5807), PY79-KO2 (NRS5808), PY79-KO3 (NRS5809), PY79-KO4 (NRS5810), PY79-KO5 (NRS5811), PY79-KO6 (NRS5812) and PY79-KO7 (NRS5787). n=1. Arrow 1 = unprocessed TapA, Arrow 2 = TasA/TapA-dependent band and Arrow 3 = low molecular weight TapA.

3.1.13. 3610 *comI* is a suitable system for studying the role of TapA processing in biofilm development

As shown above, TapA processing is dependent on secreted proteases in strain PY79, and TapA processing is conserved in several *B. subtilis* species. Therefore, we wanted to test the role of TapA processing on biofilm formation. To investigate this, a protease-free version of 3610 was required. However, the 3610 strain used in this study is not genetically competent (Nijland et al., 2010). To facilitate the mutagenesis process a competent derivative of 3610, named 3610 *comI* was obtained (Konkol et al., 2013). First it was confirmed if strain 3610 *comI* developed mature biofilms on MSgg media for 48 h. The 3610 *comI* strain formed biofilms which were indistinguishable from those formed by the 3610 strain used throughout this work (**Figure 19**) and it was judged to be an

appropriate system for understanding the role of TapA processing in biofilm formation.

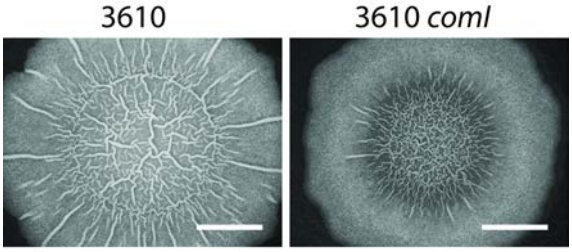


Figure 19: The biofilm morphology of 3610 and 3610 *comI*. The biofilm morphology of 3610 and 3610 *comI* (NRS6017) grown on MSgg agar for 48 h before image capture. Scale bar = 0.5 cm. Images are representative of at least 3 independent replicates.

3.1.14. An exoprotease-free version of 3610 *comI* (NRS6362) was generated

To allow for the investigation of the role exoproteases play in TapA processing and biofilm formation, an exoprotease-free 3610 *comI* strain was generated. To make this strain, 3610 *comI* was transformed using *B. subtilis* genomic DNA containing a resistance cassette in place of the gene of interest. This was possible thanks to the use of a genome-wide deletion library which is available for *B. subtilis* (Koo et al., 2017a). Through this method, an antibiotic resistance cassette was incorporated in place of the wild-type gene by homologous recombination. On confirmation of successful disruption the cassette was then excised to allow the next gene to be deleted. Gene deletions that were proving problematic to construct by this method were successfully obtained by using phage transduction. The strains generated by this method are shown in Table 11. Importantly, the exoprotease genes were deleted in the opposite order from which they were deleted in generating the PY79-KO7. In other words, in generating PY79-KO7, the *nprE* gene was disrupted first and the *bpr* was disrupted last (Table 10). In contrast, the *bpr* gene was disrupted first and the

nprE was disrupted last in generating 3610-KO7 (Table **11**). This decision was based on the above observation that PY79-KO6 and PY79-KO7 cell lysates displayed the most noticeable differences in TapA processing indicating that the genes *bpr* and *vpr* may have a specific role in processing TapA. Alternatively, it was suggested that the ability to cleave TapA is a redundant property shared by multiple or all of the exoproteases.

Strain	Genotype	strain number
3610 *comI*	Parental strain	NRS6017
3610-KO1	*Δbpr*	NRS6047
3610-KO2	*Δbpr Δvpr*	NRS6049
3610-KO3	*Δbpr Δvpr ΔnprB*	NRS6061
3610-KO4	*Δbpr Δvpr ΔnprB Δmpr*	NRS6063
3610-KO5	*Δbpr Δvpr ΔnprB Δmpr Δepr*	NRS6065
3610-KO6	*Δbpr Δvpr ΔnprB Δmpr Δepr ΔaprE*	NRS6341
3610-KO7	*Δbpr Δvpr ΔnprB Δmpr Δepr ΔaprE ΔnprE*	NRS6362

Table 11: A list of the 3610 *comI* exoprotease mutant strains generated in this study and used to aid the understanding of TapA processing in the biofilm.

3.1.15. The general exoprotease activity of 3610-KO7 is abolished

As was shown above the general exoprotease activity of PY79 was abolished on disruption of just 2 of the exoprotease genes, *nprE* and *aprE*. To allow a comparison, the impact on general exoprotease activity upon deletion of exoprotease genes in 3610 was tested. The exoprotease genes were deleted in reverse order from that in the PY79 mutants. The exoprotease activity of the exoprotease-free 3610 strain (3610-KO7), and all of the intermediate strains, were tested by growth on LB agar, that was supplemented with 1.5% (w/v) milk, for 18 h at 37^{o}C. It was found that the wild-type 3610 strain displayed protease activity, as determined by the presence of a halo around the colony (**Figure 20**A). The strain 3610 *degU*, was defective in the production of degradative

enzymes, and serves as a negative control in this assay and did not display detectable exoprotease activity (Verhamme et al., 2007). Interestingly, the halo around 3610 was larger than that observed around the PY79 colony grown under the same conditions, compare with **Figure 16**B. All of the intermediate strains from 3610-KO1 to 3610-KO6 showed a comparable halo to the 3610 (**Figure 20**A). This is consistent with the idea that 3610 has more general exoprotease activity than PY79. This is because it was shown previously that most of the intermediate strains of PY79, PY79-KO3 to PY79-KO7, did not display any detectable exoprotease activity (**Figure 16**B). Importantly, the 3610-KO7 exoprotease-free strain did not have observable proteolytic activity against milk proteins as judged by the absence of a halo around the colony (**Figure 20**A). Therefore, the generalised exoprotease activity of strain 3610-KO7 is abolished as determined by an ability to degrade milk proteins. In contrast, the exoprotease activity of PY79 is abolished on deletion of just 2 exoprotease genes.

3.1.16. The processing of TapA in the biofilm is dependent on the action of extracellular proteases

To assess the impact of deleting the exoprotease genes on TapA processing, an αTapA immunoblot was carried out using protein extracted from 3610 *comI* and 3610 exprotease mutant biofilms. Processing in the 3610-KO1 (*bpr*) strain was similar to that observed in the wild-type biofilm, with the observation of the characteristic 3 bands which included the presence of the low molecular weight TapA band of 16 kDa (**Figure 20**B, Arrow 1). The low molecular weight TapA band was no longer observed in the 3610-KO2 to 3610-KO6 intermediate strains or the exoprotease-free strain 3610-KO7 (**Figure 20**B). The loss of this 16 kDa band therefore coincided with the deletion of the *vpr* gene (the genotype

of 3610-KO2 is *Δbpr Δvpr*). This is in agreement with the result shown in **Figure 18**B in which the presence of this band was shown to be dependent on the presence of the *vpr* gene in the PY79 strain background. All strains exhibited a band resembling the TasA/TapA dependent band present in wild-type 3610 *comI* biofilms of approximately 18 kDa (**Figure 20**B, Arrow 2). Interestingly, a band of >25 kDa was visible in the 3610-KO6 and 3610-KO7 strains (**Figure 20**B, Arrow 3). This band is found below the non-specific TasA band, which was found in all of the strains (**Figure 20**B, Arrow denoted TasA). It is predicted that this band contains unprocessed TapA.

3.1.17. Exoproteases are not needed for complex colony development in 3610 biofilms

It was shown that TapA processing in 3610 biofilms is impacted by the absence of exoproteases. To determine if aberrant processing of TapA had an impact on biofilm formation the architecture was examined. Interestingly, all of the biofilms formed by these strains displayed wild-type colony complexity (**Figure 20**C). This result demonstrated that gross biofilm morphology is not impacted by the loss of the 7 exoprotease encoding genes.

3.1.18. Cell-free 3610-KO7 supernatant retains residual activity against TapA *in vitro*

It was shown earlier that *in vitro* TapA cleavage by PY79 supernatant was due to the action of exoproteases. On construction of the 3610 exoprotease mutants (**Table 11**), it was tested whether the ability of 3610 supernatant to cleave TapA *in vitro* was impacted. To test this, purified $TapA_{44\text{-}253}$ protein was incubated with spent culture supernatant from 3610-KO7 (NRS6362) and the integrity of the

protein was analysed by SDS-PAGE (**Figure 20**D). The 3610-KO7 culture supernatant retained residual proteolytic activity against TapA when compared to wild-type 3610 supernatant or the $TapA_{44-253}$ input used in the assay (**Figure 20**D). The proteolytic activity of the supernatant was abolished upon heat-treatment as before (3610 +H and 3610-KO7 +H). It was interesting to note that higher molecular weight forms of TapA were observed in the input sample and the TapA which was incubated with heat-inactivated supernatants, however, these were absent in the samples in which TapA degradation could be observed.

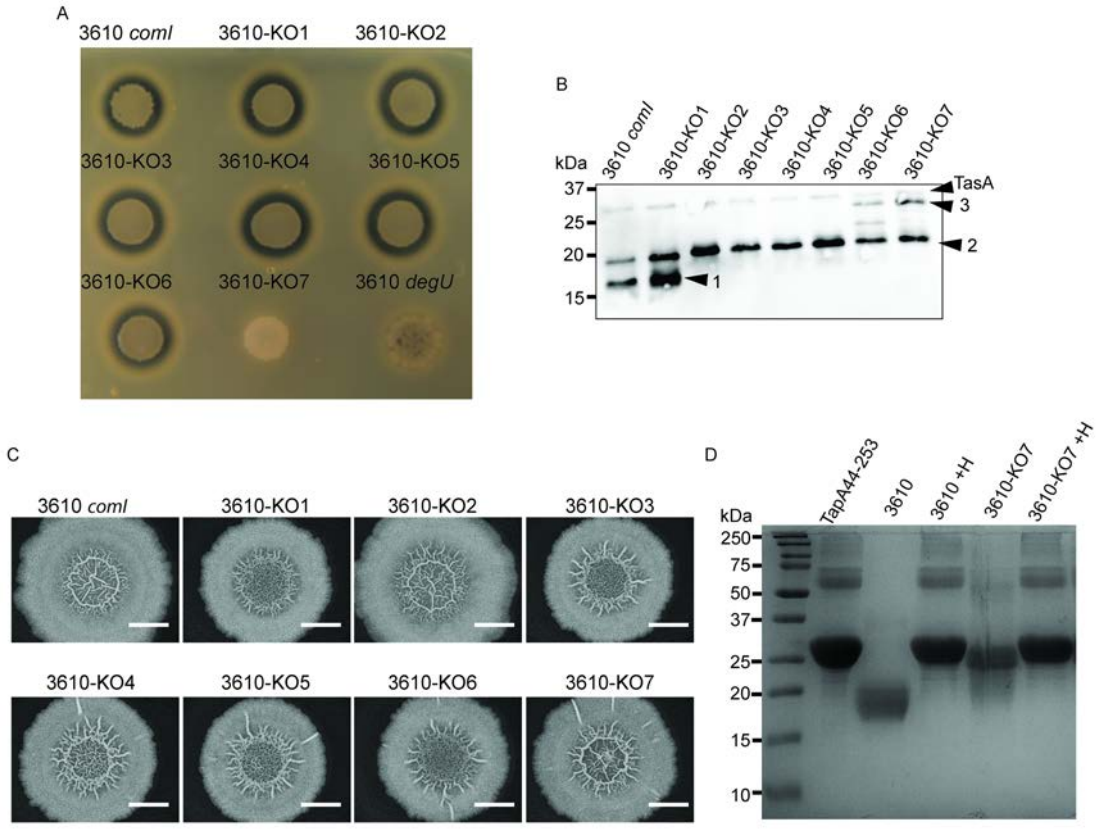


Figure 20: Biofilm formation in 3610 is independent of exoprotease activity. (A) 3610 *comI* strains were examined for protease activity by growth on LB 1.5% milk agar plates: 3610 *comI* (NRS6017), 3610-KO1 (NRS6047), 3610-KO2 (NRS6049), 3610-KO3 (NRS6061), 3610-KO4 (NRS6063), 3610-KO5 (NRS6065), 3610-KO6 (NRS6341), 3610-KO7 (NRS6362). n=3. **(B)** Protein was extracted from 3610 exoprotease mutant biofilms and analysed by αTapA immunoblot. n=2. Arrow 1 = low molecular weight TapA, Arrow 2 = TasA/TapA-dependent band and Arrow 3 = unprocessed TapA. Arrow TasA denotes the non-specific TasA band detected using the αTapA antibody. **(C)** The biofilm morphology of 3610 *comI* and exoprotease mutants

derived from this parental strain. Strains were grown on MSgg for 48 h before imaging. n=5. **(D)** $TapA_{44-253}$ protein was mixed 1:1 (vol:vol) with cell-free supernatant obtained from 3610-KO7, and incubated for 24 h at 37^{o}C, the amount of TapA protein in the assay was 150 µg. 30 µg pf protein was run on 14% SDS-PAGE and stained with InstantBlue. To heat-inactivate (+H), supernatant was incubated at 100^{o}C for 10 min. n=1. Note that this figure is an extended version of that shown in **Figure 15**B. Scale bar = 0.5 cm.

In combination, these results demonstrate that the exoproteases of strain 3610 are responsible for most of the proteolysis of TapA observed in the supernatant assay carried out. However, it may be that sufficient TapA processing occurs in 3610-KO7, based on the residual proteolytic activity displayed by 3610-KO7 supernatant, to produce a functional form of TapA needed for biofilm formation.

3.1.19. Summary

The work carried out here investigated the function of TapA in the formation of *B. subtilis* biofilms. As previously published, when either the *tapA* or *tasA* gene is deleted then biofilm complexity is lost. In agreement with published findings, reduced levels of TasA are observed in biofilms formed by the Δ*tapA* mutant. For this reason a Δ*tapA* mutant can be thought of as somewhat akin to a Δ*tapA* Δ*tasA* double mutant.

With the aid of an αTapA antibody it was found that TapA undergoes processing to a low molecular weight form *in vivo*. PY79 was used as a model system for studying the role of exoproteases in this phenomenon. It was found that TapA processing *in vivo* is dependent, at least in part, on exoprotease genes. The generation of low molecular weight TapA (16kDa) was shown to be dependent on the presence of the *vpr* gene. However, as PY79 does not form complex biofilms this work was recapitulated in 3610.

An exoprotease-free 3610 (3610-KO7) strain was developed using a competent variant of 3610 called 3610 *comI*. The generalised exoprotease activity of 3610-KO7 was abolished as judged by the ability to degrade milk proteins present in the growth substrate. Finally, it was found that TapA processing was aberrant in the 3610-KO7 mutant but that this strain still formed mature biofilms resembling wild-type in appearance. This result may be explained by the subsequent finding that 3610-KO7 supernatant has residual proteolytic activity against recombinant TapA *in vitro*. Additionally, it was shown that the low molecular weight form of TapA was dependent on the presence of the *vpr* gene, recapitulating the result shown by analysis of strain PY79.

3.2. Development of a model to probe the interaction between TapA and TasA

3.2.1. Background

We have shown, in section **3.1**, that TapA is a critical component in *B. subtilis* biofilms. In agreement with published findings, it was also shown that TapA is needed to maintain wild-type levels of TasA in the biofilm. The current model for TapA function is that it is a component of TasA fibres which extend from the cell surface, anchoring them to the cell envelope. However, there is limited knowledge as to which components of the TapA protein are needed for its function in aiding biofilm development.

To examine, in more detail, the molecular mechanisms of the interaction between TapA and TasA an assay was designed. This was based on the finding that recombinant TasA fibres are functional *in vivo*. TasA produced heterologously in *E. coli,* produces fibres *in vitro* (hereafter referred to as fTasA for fibrous TasA) (Erskine et al., 2018). This ability of fibrous TasA to form in the absence of TapA was an interesting observation. Furthermore, contrary to published work when these fibres were added exogeneously, wild-type morphology could not be restored to Δ*tasA* biofilms (Erskine et al., 2018). Additionally, when native fibres (ntTasA), extracted from 3610, were added exogeneously to Δ*tasA* biofilms there was also no noticeable increase in rugosity (Erskine et al., 2018).

Serendipitously, an additional loss of function mutation in the gene encoding *sinR* was detected in a separate Δ*tasA* isolate. The inclusion of a small region of duplication in the *sinR* coding region, essentially disrupting the coding region,

was confirmed by whole genome sequencing. The *sinR* gene is proximal to the *tapA* operon on the chromosome and so may have been disrupted during the homologous recombination event in which *tasA* was deleted. The architecture of this region of the genome is shown in **Figure 21**A. By comparison with 3610 biofilms, a *sinIR::spec* (NRS2012) mutant forms biofilms with a densely wrinkled appearance (**Figure 21**B) as first observed by (Kearns et al., 2005). In contrast, the biofilms formed by Δ*tasA sinR* (NRS5248) are flat and featureless (**Figure 21**B). Interestingly, it was found that the exogeneous provision of 20 µg of fTasA to the double mutant Δ*tasA sinR* (NRS5248) resulted in a densely wrinkled morphology, resembling that of a *sinIR::spec* mutant (**Figure 21**B). These results demonstrated the biological activity of recombinant fTasA *in vivo.* This finding is attributed to Elliot Erskine and has been published (Erskine et al., 2018).

It is not clear why an additional mutation was required to acheive exogeneous complementation of Δ*tasA* biofilms with fTasA. It is known that SinR functions as a repressor of biofilm formation inhibiting transcripton of the *tapA* and *eps* operons (Chu et al., 2006; Kearns et al., 2005), so it may be that increased levels of TapA or EPS are facilitating the function of fTasA in this assay.

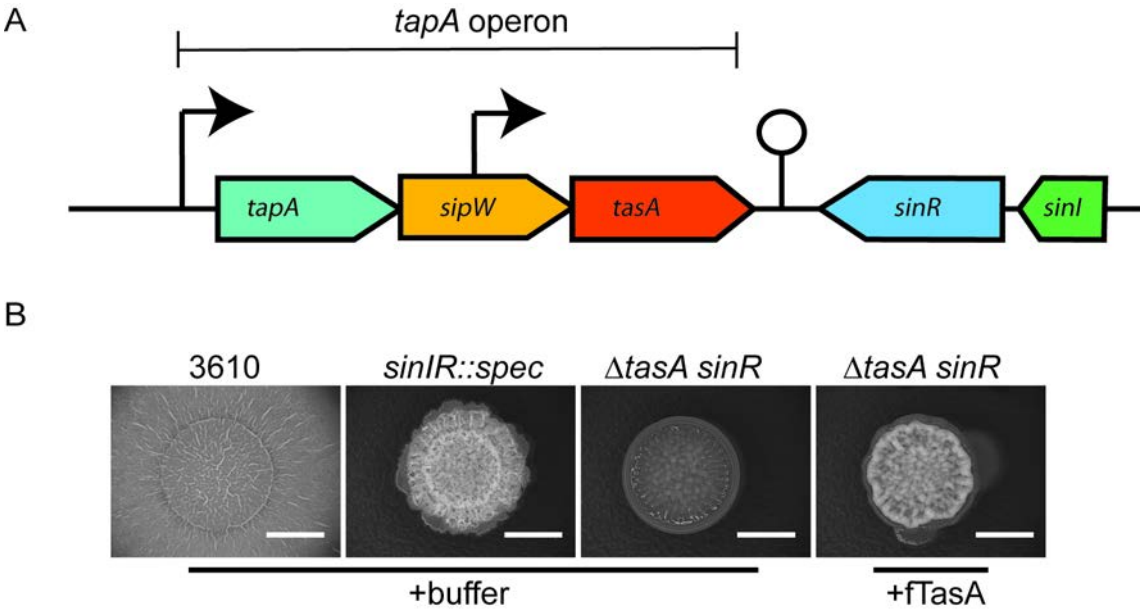


Figure 21: The exogeneous provison of fTasA restores structure to a Δ*tasA sinR* mutant. **(A)** The architecture of the *tapA* operon and the location of genes *sinR* and *sinI*. Open reading frames (ORF) are shown as coloured, elongated arrows. The bents arrows indicate the promoter which drives expression of the operon as well as the internal promoter which is found within the region of the *sipW* ORF which serves as an additional control of *tasA* gene expression. The transcriptional terminator for the *tapA* operon is shown as a stick and circle. The *sinR* and *sinI* ORFs are orientated in the opposite direction to the *tapA* operon and encoded on the opposite strand of DNA. **(B)** Biofilm images of 3610, *sinIR::spec* (NRS2012) and Δ*tasA sinR* (NRS5248). Biofilms were either spotted onto MSgg plates with buffer (+buffer) or fTasA (+fTasA). Images are representative of more than 3 independent replicates. Scale bars = 0.5 cm.

Given the result presented above, it was hypothesised that by deleting *tapA* in the strain Δ*tasA sinR,* to generate Δ*tapA* Δ*tasA sinR*, that this may provide an opportunity to study the molecular basis of the relationship between TasA and TapA. Based on published research (Romero et al., 2011, 2014), it was anticipated that without the presence of TapA then exogeneous TasA fibres would not be able to rescue biofilm formation. TapA is reported to be a component of TasA fibres which anchors them to the cell. In this way, *tapA* variants could be introduced into these strains and assessed for the ability to restore wild-type biofilm morphology when fTasA is added exogeneously. It was

hoped that this might lead to the identification of protein-protein interaction sites between TasA and TapA. However, the *tapA* gene first had to be deleted in the Δ*tasA sinR* strain background to generate the strain Δ*tapA* Δ*tasA sinR* (NRS5749).

3.2.2. TapA is not needed for the exogeneous complementation of Δ*tasA sinR* biofilms

Strain Δ*tapA* Δ*tasA sinR* (NRS5749) was generated from the parental strain ∆*tasA sinR* (NRS5248) by the deletion of the *tapA* gene using the pMAD plasmid (pNW685). After construction, since Δ*tapA* Δ*tasA sinR* had undergone multiple rounds of mutagenesis the intergrity of the strain was tested. To determine if the biofilm phenotype of the strain could be genetically complemented, the *tapA* operon was introduced to the *amyE* locus under the control of the IPTG inducible P_{spank} promoter. This generated the strain Δ*tapA* Δ*tasA sinR* + *pIPTG-tapA-sipW-tasA* (NRS5763). 3610 biofilms display characteristic corrugations, in contrast, the biofilms formed by the Δ*tasA sinR* double mutant and the triple mutant Δ*tapA* Δ*tasA sinR* displayed flat, featureless phenotypes (**Figure 22**A). It was shown that inducing expression of the *tapA* operon in the strain ∆*tapA* ∆*tasA sinR* + *pIPTG-tapA-sipW-tasA* (NRS5762) restored the biofilm to the densley wrinkled appearance which is displayed by the *sinIR::spec* biofilm (**Figure 22**A). Note that in our experience there is basal levels of expression from the P_{spank} promoter in the absence of IPTG. This explains the rugose morphology of ∆*tapA* ∆*tasA sinR* + *pIPTG-tapA-sipW-tasA* grown without IPTG (**Figure 22**A). Together these results demonstrate that the strain could be genetically complemented with respect to biofilm formation and is therefore suitable for use in the development of the assay described above.

It was predicted that addition of fTasA to Δ*tapA* Δ*tasA sinR* cells would not restore biofilm formation due to the absence of *tapA*. This would then allow for *tapA* variants to be introduced and the formation of wrinkles observed.

To test this, either 20 µg of fTasA (+fTasA) or buffer (+buffer) was added to cells before spotting onto biofilm forming media. 3610 and *sinIR::spec* (NRS2012) biofilms are shown for comparison and demonstrate that the addition of buffer or fTasA does not disrupt biofilm formation. The exogeneous provision of 20 µg of recombinant fTasA to Δ*tasA sinR* (NRS5248) cells re-instated rugosity as shown previously (**Figure 22**B). Interestingly, it was found that rugosity could be restored to the Δ*tapA* Δ*tasA sinR* (NRS5749) mutant with the addition of fTasA (**Figure 22**B). In both cases this is shown by comparison with buffer only controls where purification buffer is added to cells before spotting (**Figure 22**B). The result that fTasA was shown to be functional within the Δ*tapA* Δ*tasA sinR* biofilm was unexpected, given that TapA is thought to be needed for the anchoring of TasA fibres to the cell wall of matrix embedded 3610 cells (Romero et al., 2011). However, it may be that fTasA can still attach to the cell surface by some other means in the system tested here. Given the result that TapA was not required for fTasA activity *in vivo*, the model proposed was no longer pursued as a viable method to study TasA/TapA interaction. These results have been published (Erskine et al., 2018).

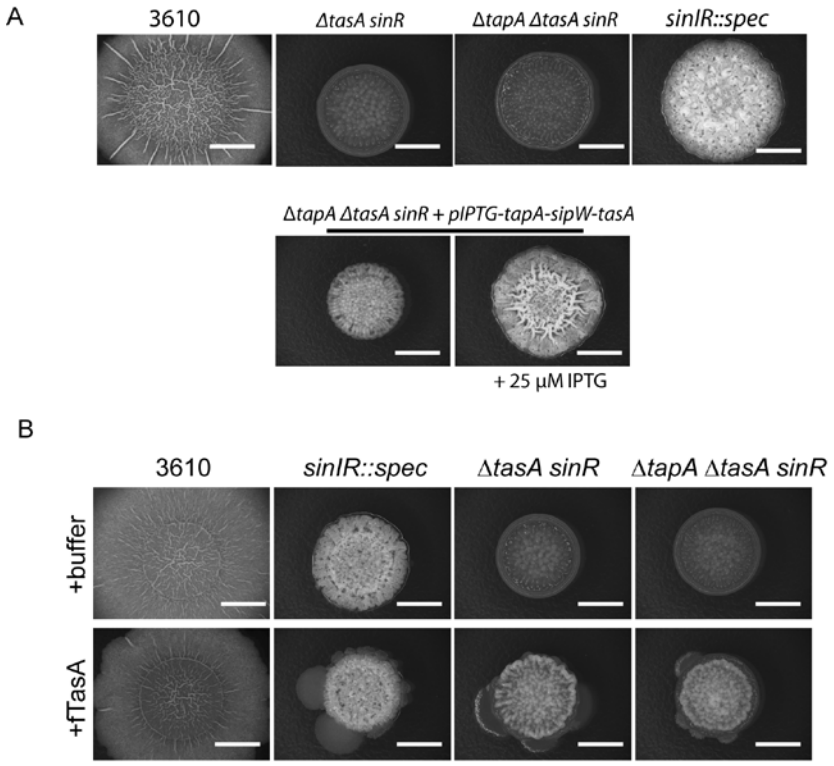


Figure 22: TapA is not needed for the function of exogeneously added fTasA in the biofilm. (A) Genetic complementation of Δ*tapA* Δ*tasA sinR* biofilms with the *tapA* operon expressed from an ectopic locus under the control of an IPTG inducible P_{spank} promoter in the presence of 25 µM IPTG (+IPTG). 3610, Δ*tasA sinR* (NRS5248), Δ*tapA* Δ*tasA sinR* (NRS5749), *sinIR::spec* (NRS2012) and Δ*tapA* Δ*tasA sinR* + *pIPTG-tapA-sipW-tasA* (NRS5763) biofilms are shown. The image of NRS5248 originates from a separate biological replicate. **(B)** *Ex vivo* addition of purification buffer (+buffer) as a control or 20 µg of fibrous TasA (+fTasA) to 3610, *sinIR::spec* (NRS2012), *ΔtasA sinR* (NRS5248) and Δ*tapA ΔtasA sinR* (NRS5749) biofilms. Scale bars = 0.5 cm.

3.3. A genetic approach to understand TapA function

Background

In the previous section it was shown that attempts to develop a model system for the study of TapA and TasA interaction were unsuccessful. The finding that TapA is not required for the bioactivity of TasA fibres when supplied *ex vivo,* suggested that a re-evaluation of TapA function and activity was warranted. Here a genetic approach is taken in an effort to understand the function of TapA.

Previously, the TRUST algorithm has been used to search for imperfect repeats within the TapA protein sequence as a way to identify functionally important sequence regions (Romero et al., 2014). The presence of internal repeats within coding sequences can be indicative of duplication events which can be important for the evolution of new or modified protein function (Szklarczyk and Heringa, 2004). This analysis resulted in the detection of an internal repeat sequence: i) from amino acids 50 to 68 and ii) from 194 to 237. It was found that deletion of the region from 194 to 237 had no impact on TapA function, within the context of biofilm formation. However, removing amino acids 50-57 resulted in a TapA protein ($TapA^{\Delta 50-57}$) that could no longer restore wild-type biofilm formation when produced in a Δ*tapA* mutant background. It was found that this protein acted in a dominant negative manner when it was produced in wild-type 3610 cells that had the wild-type TapA protein (Romero et al., 2014). This resulted in defective biofilm structure at early stages of formation, from 12-18 h, and at 24 h. Replacement of residues 50-57 with the amino acid sequence LGPGIGNG also resulted in functionless TapA. It was found that cells

harbouring only the TapA$^{\Delta 50\text{-}57}$ form of the protein had a defect in TasA localisation which was now only found in the cell portion of fractionated biofilms instead of both the cell and matrix fractions. Immunocytochemistry with αTapA antibodies, in combination with microscopy was used to detect the localization of TapA on *B. subtilis* cells. By this method TapA was found to organize in discrete foci on the cell surface. It was also found that purified TapA protein assisted TasA fibre polymerization *in vitro*. Interestingly, TapA$^{\Delta 50\text{-}57}$ failed to assist in polymerisation when studied in the same assay (Romero et al., 2014).

The aim here was to take an independent genetic approach to determine sequence features of *tapA* that are needed for TapA function. Natural variants of *tapA* found in related *Bacillus* species initially served as a source of genetic diversity for the interrogation of the *B. subtilis* sequence. Protein sequence comparisons illuminated highly conserved regions and redundant regions which became the focus of further investigation. This ultimately resulted in the identification of the minimal functional genetic unit of *tapA* which was found to encode the first 57 amino acids of the protein (TapA$_{1\text{-}57}$). Finally, the requirement of a signal sequence for the activity of TapA in the biofilm, as predicted *in silico*, was tested experimentally.

3.3.1. The *tapA* coding sequence from related *Bacillus* species can restore biofilm architecture to 3610 Δ*tapA* biofilms

As a starting point for investigating important features of TapA, the biofilm activity of TapA variants from related *Bacillus* species were analysed. The ability of *B. pumilis (tapA$_{B_pum}$)*, *B amyloliquefaciens* (*tapA$_{B_amy}$*) and *B. paralicheniformis* (*tapA$_{B_para}$*) *tapA* ORFs to restore biofilm formation in 3610 Δ*tapA* (NRS3936) was tested. The *tapA* operon from each of these species is very well conserved, with the *tapA* gene being the most variable ORF within the operon (Appendix: **Figures S1-S5** and **Table S1**). The *tapA* homologs were integrated at the *amyE* locus under the control of an IPTG inducible promoter. The strains were grown on MSgg media, in the presence of IPTG, for 48 h to promote biofilm formation. As shown previously, the biofilm formed by the parental 3610 strain displayed a wrinkled morphology (**Figure 23**A) and the Δ*tapA* (NRS3936) biofilm exhibited a smooth surface (**Figure 23**A). Biofilm complexity was restored to the *ΔtapA* + *pIPTG-tapA* (NRS5045) strain when the *B. subtilis tapA* variant was induced by the presence of IPTG. Heterologous expression of the *tapA* homologs was shown to restore biofilm architecture to Δ*tapA* + *pIPTG-tapA$_{B_pum}$*, Δ*tapA* + *pIPTG-tapA$_{B_amy}$* and Δ*tapA* + *pIPTG-tapA$_{B_para}$* biofilms (**Figure 23**A). This result demonstrated that *tapA* homologs from related *Bacillus* species encode TapA homologs which are functional in the *B. subtilis* 3610 Δ*tapA* strain. To examine important sequence features in common between the TapA homologs the protein sequences were analysed bioinformatically.

3.3.2. The conserved region 111-SKWKWELHK-120 is not needed for TapA function in the biofilm

A clustal omega protein sequence alignment of TapA homologs is shown in **Figure 23**B. It was observed that there is a well conserved region between residues 111-120 (**Figure 23**B, Red box). This suggested that the conservation of this region in the functional TapA homologs tested, may be indicative of an important role for this region in conveying biological activity to TapA in the biofilm. Therefore, a *tapA* variant which encoded full-length TapA with residues 113-118 changed to alanine was constructed. The variant $TapA_{111\text{-}SKAAAAAAKL\text{-}120}$ is hereafter called $TapA_{\Delta cons}$ for convenience. The gene encoding the $TapA_{\Delta cons}$ variant was introduced at the *amyE* locus under the control of an IPTG-inducible promoter to make the strain Δ*tapA* + *pIPTG-tapA*$_{\Delta cons}$ (NRS5794). To test the functionality, the strain was grown on MSgg media for 48 h to promote biofilm formation. The biofilm formed by the parental 3610 strain exhibited complex, wrinkled morphology (**Figure 23**C) and as expected the Δ*tapA* (NRS3936) biofilm was flat and featureless (**Figure 23**C). The induction of *tapA* in strain Δ*tapA* + *pIPTG-tapA* (NRS5045) restored wild-type biofilm architecture. Interestingly, growing the strain Δ*tapA* + *pIPTG-tapA*$_{\Delta cons}$ (NRS5794) in the presence of IPTG also restored biofilm complexity (**Figure 23**C). This result indicated that this conserved region is not required for TapA activity in the biofilm.

To observe the $TapA_{\Delta cons}$ variant *in vivo*, biofilm lysates were analysed by αTapA immunoblot. The stained SDS-PAGE, which serves as a loading control for this experiment, is shown in **Figure 23**D. As discussed earlier, the αTapA antibody recognises 3 bands with the lowest molecular weight band being confirmed as TapA and the 2 higher bands being TasA-dependent. Only the low

molecular weight TapA band was observed in the Δ*tasA* sample (**Figure 23**E). As before, the 3 bands were observed in the 3610 biofilm lysates but absent in the Δ*tapA* lysate (**Figure 23**E). The 3 bands could be observed in lysates originating from the strain Δ*tapA* + *pIPTG-tapA* (NRS5045) in an IPTG-dependent manner. Interestingly, the lower 2 bands were absent in the Δ*tapA* + *pIPTG-tapA*$_{\Delta cons}$ (NRS5794) lysate even in the presence of IPTG (**Figure 23**E). However, the TasA-dependent band at ~30 kDa was present on induction with IPTG, moreover, the biofilm presented a wild-type phenotype. This result was unexpected given that the αTapA antibody detected TasA (non-specifically) yet was unable to detect the TapA$_{\Delta cons}$ variant. Taken together these results indicated that this region is important for antibody recognition but that it was not required for biofilm formation. Next, the question was raised as to whether there are distinguishable features between the *tapA* homologs that may help elucidate the functional form of the protein.

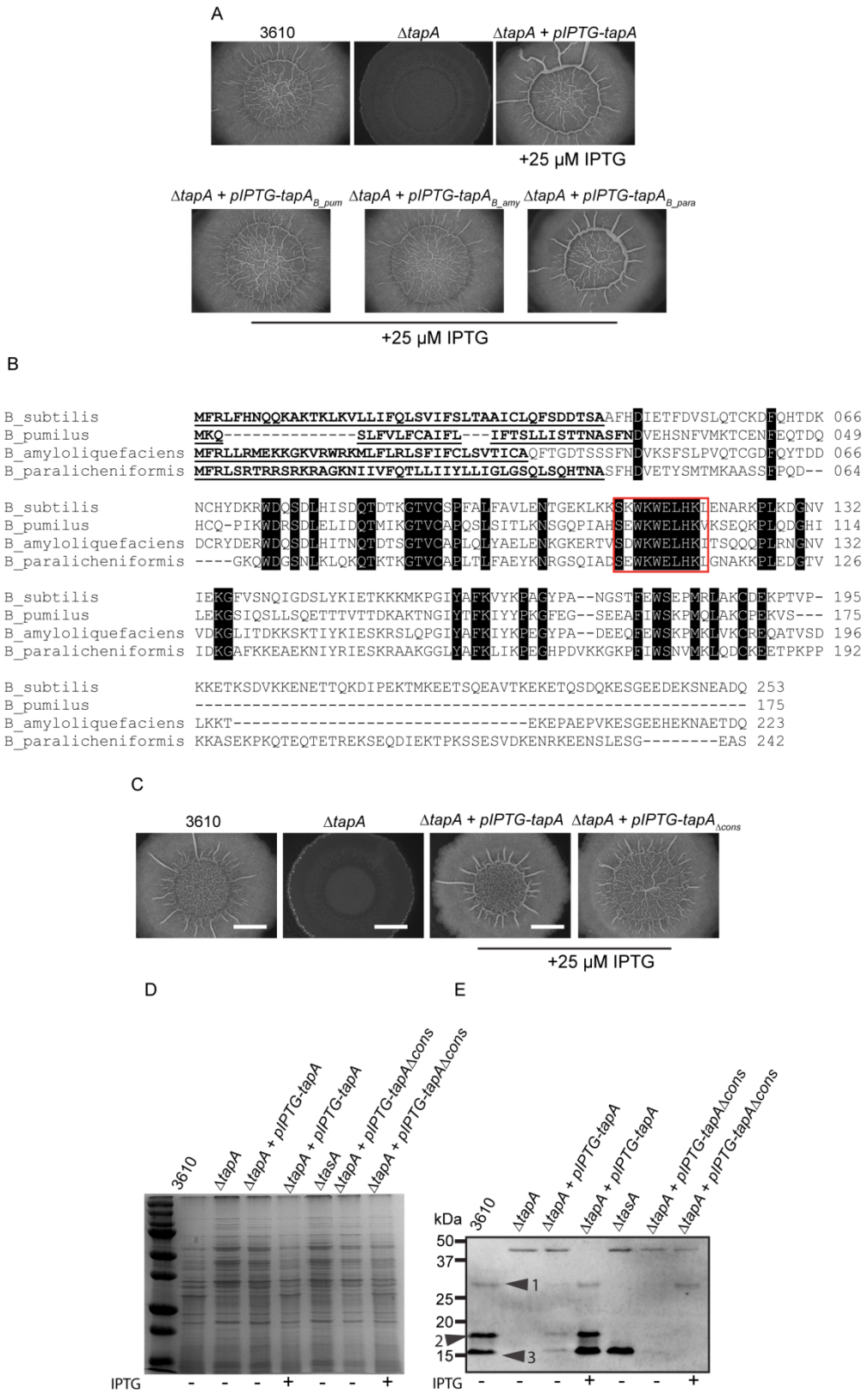


Figure 23: Homologous TapA proteins from related Bacillus species display acitvity in the 3610 biofilm. (A) The morphology of Δ*tapA* biofilms with the induction of *tapA* homologs: 3610, Δ*tapA* (NRS3936), Δ*tapA* + *pIPTG*-tapA (NRS5045), Δ*tapA* + *pIPTG-tapA*$_{B\_pum}$ (NRS5046), Δ*tapA* + *pIPTG-tapA*$_{B\_amy}$ (NRS5047), Δ*tapA* + *pIPTG-tapA*$_{B\_para}$ (NRS5743). n>3. **(B)** A clustal omega alignment of TapA protein sequences derived from *Bacillus subtilis* and related *Bacillus* species. Highlighted in black are residues with 100% conservation between the

4 sequences. A red box is used to demark the conserved region from amino acids 111-120. **(C)** Biofilms of 3610, Δ*tapA*, Δ*tapA* + *pIPTG-tapA* (NRS5045) and Δ*tapA* + *pIPTG-tapA*$_{\Delta cons}$ (NRS5794) grown in the presence of 25 µM IPTG for 48 h. n=3. **(D)** Biofilm lysates of the same set of strains as shown in **(C)** analysed by SDS-PAGE on a 12% acrylamide gel, 72 µg of protein was loaded for each sample, this gel stained with InstantBlue serves as a loading control for **(E)**. n=3. **(E)** αTapA immunoblot of biofilm lysates. n=1 **(E)**. - = 0 µM IPTG + = 25 µM IPTG. Arrow 1 corresponds to TasA (30 kDa band), arrow 2 indicates a TasA/TapA-dependent band (~18 kDa) and arrow 3 is TapA (16 kDa). Scale bars = 0.5 cm.

3.3.3. The minimal genetic unit required for TapA function encodes the first 57 amino acids of the protein

Revisiting the TapA sequence alignment it was found that *B. pumilis* TapA, at only 175 amino acids, is the shortest of the homologs tested in this study. Strikingly, the TapA homolog from *B. pumilis* was shown to be functional in *B. subtilis*. Given this result, the aim of the next set of experiments was to determine whether the C-terminus of TapA is necessary for biofilm function. To test this, a *B. subtilis tapA* variant was constructed encoding amino acids 1-193 (TapA$_{B_sub1\text{-}193}$). This construct is designed to mimic the truncated C-terminus observed in *B. pumilis* TapA. Biofilm formation was analysed by growth of the bacterial strains on MSgg media for 48 h before image capture. The 3610 biofilm exhibited a typical rugose, complex appearance while the Δ*tapA* strain was featureless and smooth (**Figure 24**A). The Δ*tapA* + *pIPTG-tapA* (NRS5045) and Δ*tapA* + *pIPTG-tapA*$_{1\text{-}193}$ (NRS5744) strains were found to form wild-type biofilms when expression was induced with IPTG (**Figure 24**A). This result demostrated that *tapA*$_{1\text{-}193}$ encodes a functional product and that the portion of the C-terminus from amino acids 193-253 is not needed for activity in the context of biofilm formation. This is consistent with a previous report that amino acids 194 to 237 can be removed without impacting TapA function (Romero et al., 2014).

The result of this experiment, in combination with th role of exoproteases in TapA processing, raised the question of how much of the C-terminus is required for its activity in the biofilm. To address this, a series of truncated forms of the *tapA* ORF were designed to determine the minimal coding region that was needed to yield an active gene product (termed the minimal functional unit: mfu). Through the systematic analysis of 22 constructs the Δ*tapA* biofilm defect was found to be rescued upon provision of a very short region of the gene encoding amino acids 1-57. This was demonstrated with the strain Δ*tapA* + *pIPTG-tapA*$_{1-57}$ (NRS6041) forming wild-type biofilms (**Figure 24**A). In this case biofilm formation was considered complemented when rugose structure was restored to the colony in the presence of IPTG. In fact, all of the constructs encoding TapA variants from amino acids 57 up to 253 in length restored complexity to the biofilm (**Figure 24**A). In contrast, the biofilms formed by the strain Δ*tapA* + *pIPTG-tapA*$_{1-56}$ (NRS6025) displayed only partial structuring. Moreover, increasing the IPTG concentration, from 25 µM to 1 mM, to boost transcription of the coding region, did not restore wild-type morphology to Δ*tapA* + *pIPTG-tapA*$_{1-56}$ biofilms (**Figure 24**A). Biofilms formed by Δ*tapA* + *pIPTG-tapA*$_{1-50}$ (NRS6002) phenocopied the appearance of Δ*tapA* biofilms indicating that *tapA*$_{1-50}$ does not encode a functional form of the protein (**Figure 24**A). Taken together these results confirm that the minimal functional unit of *tapA* which is sufficient to convey biofilm activity is that which is encoded by *tapA*$_{1-57}$. It was surprising that such a limited part of the protein was needed given the level of amino acid conservation between TapA orthologues in the core of the *B. subtilis* sequence from amino acids 58-195, which is shown here to be redundant in regards to the biofilm activity of TapA. For comparison an

illustration of WT TapA (**Figure 24**B) and $TapA_{1-57}$ (**Figure 24**C) are depicted to convey the strikingly small proportion of TapA that is needed for activity.

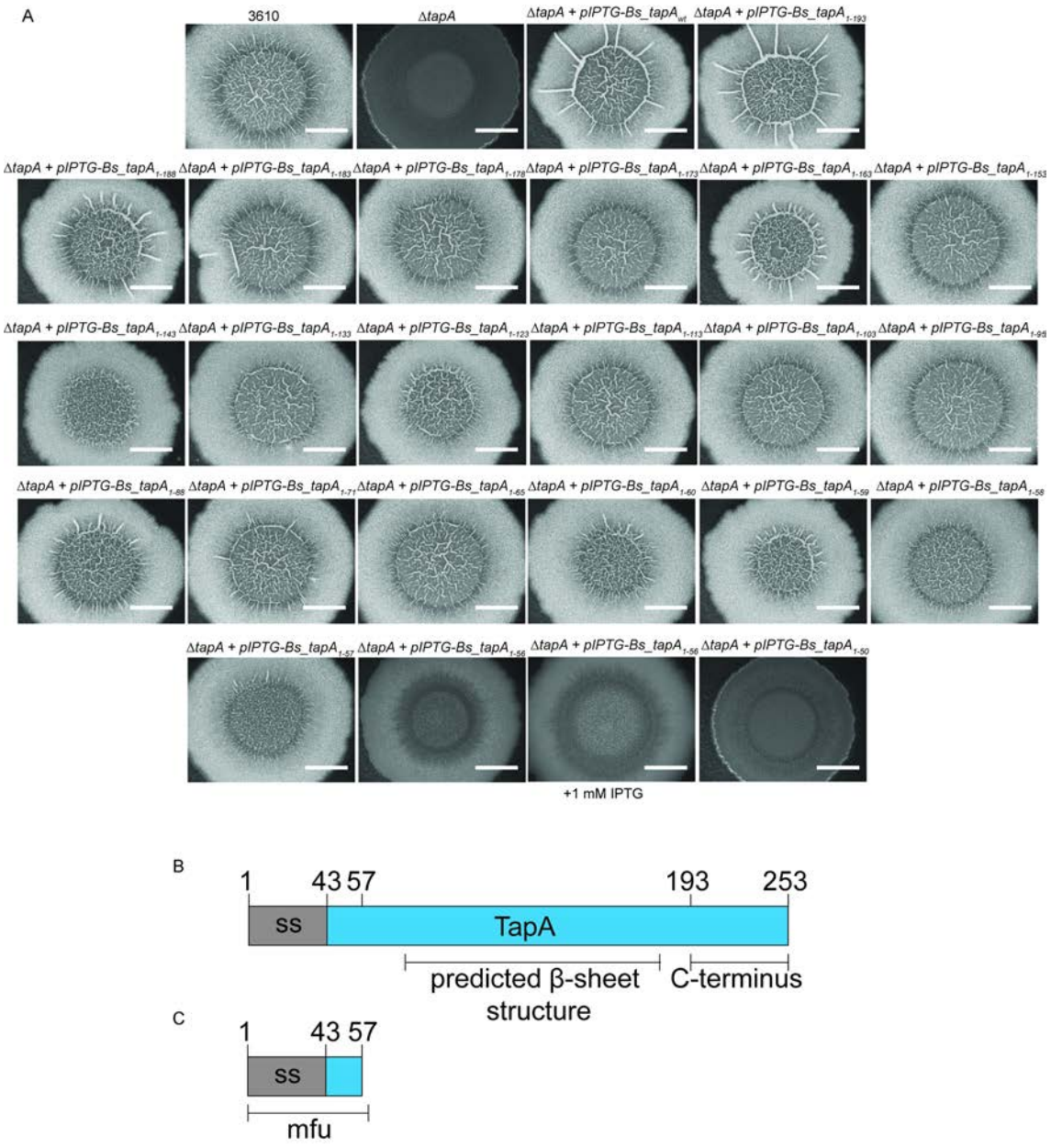


Figure 24: The minimal functional unit of *tapA* encodes the first 57 amino acids of the protein. (A) Biofilm morphology of Δ*tapA* biofilms producing truncated TapA variants with *tapA* expressed from the *amyE* locus in an IPTG-dependent manner. Strains harbouring IPTG-inducible promoters are grown as biofilms in the presence of 25 µM IPTG unless otherwise stated. The following strains are shown: 3610, Δ*tapA* (NRS3936), Δ*tapA* + *pIPTG-tapA*$_{wt}$ (NRS5045), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-193}$ (NRS5744), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-188}$ (NRS5789), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-183}$ (NRS5790), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-178}$ (NRS5791), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-163}$ (NRS5814), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-153}$

(NRS5813), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-143}$ (NRS5806), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-133}$ (NRS5805), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-123}$ (NRS5819), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-113}$ (NRS5988), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-103}$ (NRS5989), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-95}$ (NRS6005), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-88}$ (NRS6004), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-71}$ (NRS6003), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-65}$ (NRS6005), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-60}$ (NRS6044), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-59}$ (NRS6043), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-58}$ (NRS6042), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-57}$ (NRS6041), Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-56}$ (NRS6025) and Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-50}$ (NRS6002). *Bs_tapA* = *B. subtilis tapA*. Scale bar = 0.5 cm. Images representative of at least 3 independent replicates. **(B)** The full-length TapA protein and **(C)** the TapA$_{1-57}$ minimal functional unit. These schematics show the TapA protein, highlighting regions of interest such as the signal sequence (grey), the mature protein after signal cleavage (blue) and the predicted β-sheet structure found in the core of the protein.

3.3.4. Truncated forms of TapA are not detected with the αTapA antibody

It is demonstrated above that *tapA* variants encoding proteins truncated at the C-terminus are still functional, even when only the first 57 amino acids of the protein are encoded. To detect the functional component of the TapA protein an αTapA immunoblot of biofilm lysates was conducted. Biofilms were grown as described before and lysed to facilitate the extraction of protein. These extracts were then analysed by immunoblot. As before 3 bands were observed in the 3610 strain which were absent in Δ*tapA* (NRS3936) lysates but recovered on induction of *tapA* in the strain Δ*tapA* + *pIPTG-tapA* (NRS5045) (**Figure 25**A). Biofilm lysates originating from the strain Δ*tapA* + *pIPTG-tapA*$_{1-193}$ (NRS5744) gave the same banding pattern as wild-type lysate when grown in the presence of IPTG (**Figure 25**A). Low molecular weight TapA is detected in Δ*tapA* + *pIPTG-tapA*$_{1-193}$ (NRS5744) biofilms with the same apparent molecular weight (16 kDa) as is observed in wild-type lysates (**Figure 25**A). It is interesting to note that TapA$_{1-193}$ is predicted to have a molecular weight of 17.31 kDa after cleavage of a signal peptide at amino acids 1-43 to give TapA$_{44-193}$. This indicated that TapA$_{44-193}$ maybe similar to the form present *in vivo* or that both

TapA and $TapA_{44-193}$ undergo further processing to produce the 16 kDa TapA band. To test this prediction, biofilm lysates of 3610, Δ*tapA* (NRS3936), Δ*tasA* (NRS5267), *ΔtapA + pIPTG-tapA$_{1-188}$, ΔtapA + pIPTG-tapA$_{1-183}$ and ΔtapA + pIPTG-tapA$_{1-178}$* were obtained and analysed by αTapA immunoblot as before. Use of the αTapA antibody led to the detection of 3 bands in the 3610 sample and these were absent in Δ*tapA* (**Figure 25**B). The low molecular weight form of TapA was not detected in the lysates of *ΔtapA + pIPTG-tapA$_{1-188}$, ΔtapA + pIPTG-tapA$_{1-183}$ or ΔtapA + pIPTG-tapA$_{1-178}$* (**Figure 25**B). This is despite biofilm architecture being restored to wild-type levels when these strains are grown in the presence of IPTG. These results support the conclusion that $TapA_{44-193}$ is similar to the low molecular weight form detected in 3610 biofilms. A band with an intermediate apparent molecular weight between the 18 kDa and 16 kDa bands was present in *ΔtapA + pIPTG-tapA$_{1-188}$* (**Figure 25**B). The 18 kDa band was shown in the first results chapter to be dependent on the presence of both *tasA* and *tapA* and is absent in both Δ*tasA* and Δ*tapA* mutants (**Figure 25**B). Therefore, it is not possible to confirm whether the band detected in *ΔtapA + pIPTG-tapA$_{1-188}$* represents an alternatively processed form of TapA or TasA. However, it can be concluded that the αTapA antibody is recognising a portion of the protein which is absent in the minimal functional forms. Additionally, the αTapA antibody still detected a band of 30 kDa representing TasA in the strains expressing genes encoding truncated forms of TapA (**Figure 25**A and B, Arrow 1). Given these findings it is likely that the low molecular weight, 16 kDa, form of TapA is stable, but not important, for the function of the protein in biofilm formation.

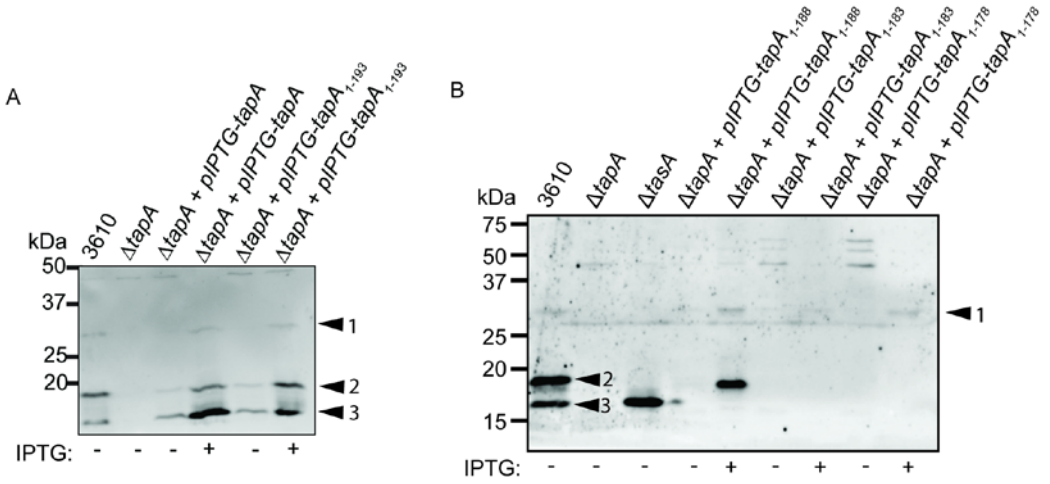


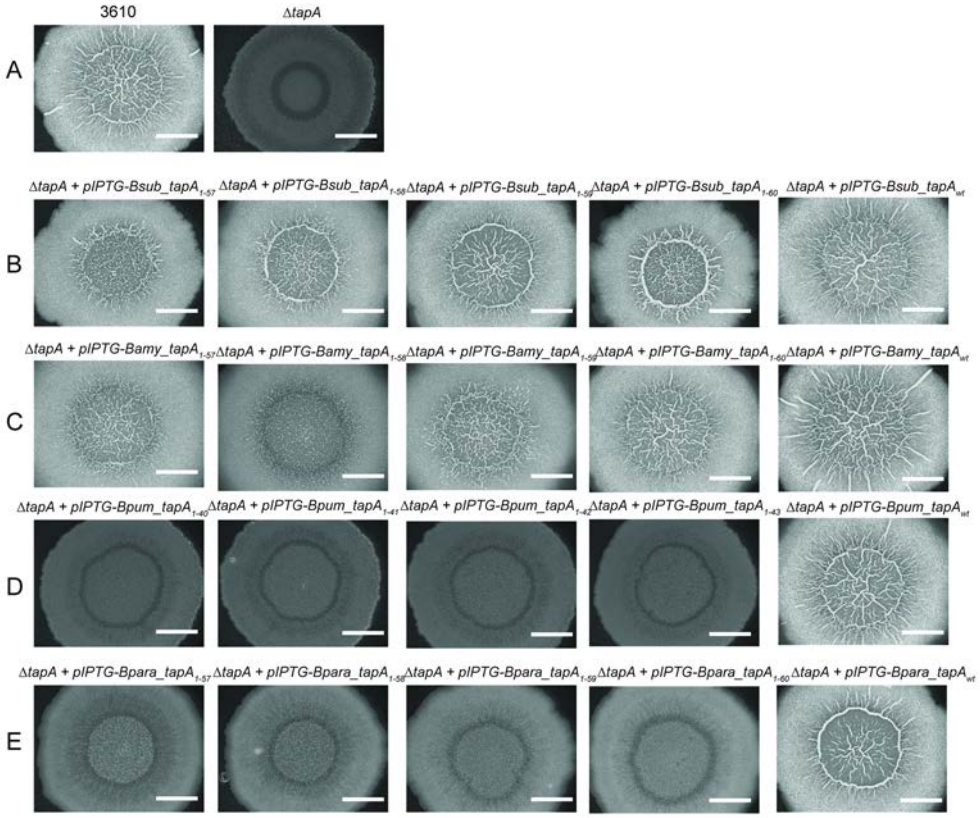
Figure 25: The αTapA antibody does not recognise the functional portion of the TapA protein. (A) αTapA immunoblot of biofilm lysates of 3610, Δ*tapA* (NRS3936), ∆*tapA* + *pIPTG-tapA*$_{wt}$ (NRS5045) and Δ*tapA* + *pIPTG-tapA-Bs_tapA*$_{1-193}$ (NRS5744). n=1 **(B)** αTapA immunoblot of biofilm lysates of the strains: 3610, Δ*tapA* (NRS3936), Δ*tasA* (NRS5267), *ΔtapA* + *pIPTG-tapA*$_{1-188}$ (NRS5789), *ΔtapA* + *pIPTG-tapA*$_{1-183}$ (NRS5790) *and ΔtapA* + *pIPTG-tapA*$_{1-178}$ (NRS5791). n=1. Arrow 1 corresponds to TasA (30 kDa band), arrow 2 indicates a TasA/TapA-dependent band (~18 kDa) and arrow 3 is TapA (16 kDa). - = 0 µM IPTG, + = 25 µM IPTG.

3.3.5. Minimal forms of TapA orthologues display a varying level of functionality

Since it was possible to identify a minimal functional form of *B. subtilis* TapA, it was postulated that determining the minimal units for the TapA orthologues may lead to important insight into key amino acid features needed for function. To address this, *B. amyloliquefaciens (Bamy)*, *B. pumillis* (*Bpum*) and *B. paralicheniformis* (*Bpara*) TapA variants were designed to mimic *B. subtilis* truncations from amino acids 1-57, 1-58, 1-59 and 1-60 (this numbering denotes the length of these orthologue truncations with relation to the *B. subtilis* TapA sequence). The genes encoding the truncated *tapA* orthologues were then integrated at the *amyE* locus of the 3610 Δ*tapA* (NRS3936) strain and expression was induced with the inclusion of IPTG in the growth substrate. Biofilms were grown for 48 h before images were captured. 3610 biofilms were observed to have a wrinkled appearance and Δ*tapA* biofilms displayed a flat

biofilm lacking the characteristic corrugations of the wild-type colony (**Figure 26**A). The Δ*tapA* + *pIPTG-Bsub_tapA*$_{WT}$ (NRS5045) strain displayed wild-type biofilm appearance and was included as a control to demonstrate the assay was functioning as expected (**Figure 26**B). As shown previously, biofilms formed by strains expressing minimal forms of the *B. subtilis tapA*: Δ*tapA* + *pIPTG-Bsub_tapA*$_{1-57}$ (NRS6041), Δ*tapA* + *pIPTG-Bsub_tapA*$_{1-58}$ (NRS6042), Δ*tapA* + *pIPTG-Bsub_tapA*$_{1-59}$ (NRS6043) and Δ*tapA* + *pIPTG-Bsub_tapA*$_{1-60}$ (NRS6044) all had the rugose surface appearance of wild-type colonies (**Figure 26**B). The Δ*tapA* + *pIPTG-Bamy_tapA*$_{1-57}$ (NRS6504) and Δ*tapA* + *pIPTG-Bamy_tapA*$_{1-58}$ (NRS6505) strains were observed to regain some level of structure when expression was induced, however these were judged qualitatively to have reduced activity when compared to WT *B. subtilis* TapA (**Figure 26**B). The Δ*tapA* + *pIPTG-Bamy_tapA*$_{1-59}$ (NRS6504) and Δ*tapA* + *pIPTG-Bamy_tapA*$_{1-60}$ (NRS6505) strains were shown to have more activity, re-instating biofilms to wild-type in appearance (**Figure 26**C). In contrast, biofilms formed by Δ*tapA* + *pIPTG-Bpum_tapA*$_{1-57}$ (NRS6508), Δ*tapA* + *pIPTG-Bpum_tapA*$_{1-58}$ (NRS6509), Δ*tapA* + *pIPTG-Bpum_tapA*$_{1-59}$ (NRS6510) and Δ*tapA* + *pIPTG-Bpum_tapA*$_{1-60}$ (NRS6511) resembled Δ*tapA* biofilms in appearance (**Figure 26**D). This indicated that these forms of *B. pumilis* TapA are either not functional or unstable. Finally, the Δ*tapA* + *pIPTG-Bpara_tapA*$_{1-57}$ (NRS6512), Δ*tapA* + *pIPTG-Bpara_tapA*$_{1-58}$ (NRS6513), Δ*tapA* + *pIPTG-Bpara_tapA*$_{1-59}$ (NRS6514) and Δ*tapA* + *pIPTG-Bpara_tapA*$_{1-60}$ (NRS6515) strains displayed biofilms in which structure was only partially re-instated, indicating that these TapA variants retained partial activity (**Figure 26**E). A summary of the phenotypes shown in **Figure 26** are presented in **Table 12**.

To summarise, a minimal functional form of *B. amyloliquefaciens tapA* could be determined as that which encodes the first 59 amino acids of the protein. However, minimal functional forms of the *B. pumilis tapA* and *B. paralicheniformis tapA* genes could not be identified using the minimal functional unit of *B. subtilis tapA* as a basis for their construction. The percentage sequence identity is highest between *B. subtilis* TapA and *B. amyloliquefaciens* TapA (48.89%) than it is between *B. subtilis* TapA and *B. pumilis* TapA or *B. paralicheniformis* TapA (41.48% and 39.11%, respectively). This may explain why *B. amyloliquefaciens* TapA displays the most functionality within the *B. subtilis* Δ*tapA* system deployed here. Furthermore, it was noted that based upon secondary structure analysis that there is a β-strand predicted to be present in the region of the protein for each of the TapA homologs from amino acids 51 to 56 (based on the numbering in *B. subtilis* TapA) (**Figure 26**F). This region is conserved most strongly between *B. subtilis* TapA and *B. amyloloiquefaciens* TapA and was therefore chosen as a target for further investigation.



F

Highlighted in grey is a predicted β-strand

```
B_subtilis          MFRLFHNQQKAKTKLKVLLIFQLSVIFSLTAAICLQFSDDTSAAFHDIETFDVSLQTCKDFQHTDK 066
B_pumilus           MKQ-------------SLFVLFCAIFL---IFTSLLISTTNASFNDVEHSNFVMKTCENFEQTDQ 049
B_amyloliquefaciens MFRLLRMEKKGKVRWRKMLFLRLSFIFCLSVTICAQFTGDTSSSFNDVKSFSLPVQTCGDFQYTDD 066
B_paralicheniformis MFRLSRTRRSRKRAGKNIIVFQTLLIIYLLIGLGSQLSQHTNASFHDVETYSMTMKAASSFPQD-- 064
```

Figure 26: Biofilm morphology of ΔtapA biofilms producing truncated version of TapA homologs demonstrate varying degrees of activity. (A) Images of wild-type 3610 and Δ*tapA* biofilms. The image of the Δ*tapA* biofilm was taken as part of a different biological repeat of the same experiment. **(B)** *B. subtilis tapA* variants encoding truncated proteins produced in a Δ*tapA* strain background. **(C)** *B. amyloliquefaciens tapA* variants encoding truncated proteins. **(D)** *B. pumilis tapA* variants encoding truncated proteins. **(E)** *B. paralicheniformis tapA* variants encoding truncated proteins. **(F)** The sequences of the TapA homologs are shown. The predicted signal sequence is highlighted in bold and underlined text and shown in grey is a conserved β-strand region which was predicted by the JPRED 4 server. The number to the right of the sequences refers to the amino acids (Drozdetskiy et al., 2015). See **Table 12** for a complete list of the strains used in this experiment with strain numbers. 25 µM IPTG was present in all cases except for 3610 and Δ*tapA* biofilms. Scale bars = 0.5 cm. For figures **(A)**-**(E)** n=3.

tapA variant[a]	Biofilm phenotype[b]	Strain number
Bsub_tapA$_{WT}$*	WT	NRS5045
Bsub_tapA$_{1\text{-}57}$*	WT	NRS6041
Bsub_tapA$_{1\text{-}58}$*	WT	NRS6042
Bsub_tapA$_{1\text{-}59}$*	WT	NRS6043
Bsub_tapA$_{1\text{-}60}$*	WT	NRS6044
Bamy_tapA$_{WT}$**	WT	NRS5047
Bamy_tapA$_{1\text{-}57}$**	partial	NRS6504
Bamy_tapA$_{1\text{-}58}$**	partial	NRS6505
Bamy_tapA$_{1\text{-}59}$**	WT	NRS6506
Bamy_tapA$_{1\text{-}60}$**	WT	NRS6507
Bpum_tapA$_{WT}$***	WT	NRS5046
Bpum_tapA$_{1\text{-}40\ (Bsub\ 1\text{-}57)}$***	defect	NRS6508
Bpum_tapA$_{1\text{-}41\ (Bsub\ 1\text{-}58)}$***	defect	NRS6509
Bpum_tapA$_{1\text{-}42\ (Bsub\ 1\text{-}59)}$***	defect	NRS6510
Bpum_tapA$_{1\text{-}43\ (Bsub\ 1\text{-}60)}$***	defect	NRS6511
Bpara_tapA$_{WT}$****	WT	NRS5741
Bpara_tapA$_{1\text{-}57}$****	partial	NRS6512
Bpara_tapA$_{1\text{-}58}$****	partial	NRS6513
Bpara_tapA$_{1\text{-}59}$****	partial	NRS6514
Bpara_tapA$_{1\text{-}60}$****	partial	NRS6515

Table 12: Qualitative descriptions of the biofilm phenotypes of orthologous TapA C-terminal truncations as determined by production in the background strain 3610 Δ*tapA* as an experimental system. [a] The numbering of amino acids is the same for all TapA variants except *B. pumilis* and the equivalent number for TapA *B. subtilis* residues is shown within parenthesis in the first column of the table. [b] Biofilm phenotypes are described qualitatively as: WT- prominent corrugations resembling WT *B. subtilis* biofilms, partial– failure to form prominent wrinkles and defect– flat, featureless biofilms resembling a Δ*tapA* mutant biofilm phenotype. *Allele of *tapA* amplified from NCIB 3610 (*B_sub*), **Allele of *tapA* amplified from *Bacillus amyloliquefaciens* strain FZB42 (*B_amy*), ***Allele of *tapA* amplified from *Bacillus pumilis* strain SAFR-032 (*B_pum*) and ****Allele of *tapA* amplified from *Bacillus paralicheniformis* (*B_para*).

3.3.6. Mutational analysis on the minimal functional unit of TapA identifies key residues for TapA function

The aim of the next set of experiments was to understand the functional components of TapA$_{1\text{-}57}$ and the role of specific amino acids in conveying function. We focussed this analysis on the region after the signal peptide cleavage point. To begin with, the residues F45 and D47 were chosen for examination as they are conserved between the *B. subtilis*, *B.*

amyloliquefaciens, *B. pumilis* and *B. paralicheniformis* TapA homologs. Next, the potential β-strand structure from amino acids 51 to 56, which was identified in the previous section, was chosen for further examination. This region is constituted by alternating polar and hydrophobic residues and is a conserved feature between the minimal functional units of *B. amyloliquefaciens* TapA and *B. subtilis* TapA. Given a β-strand conformation the hydrophobic side chains of the amino acids would be orientated in the same direction and the polar side chains pointed in the opposite direction giving rise to distinct hydrophobic and polar faces (**Figure 27**).

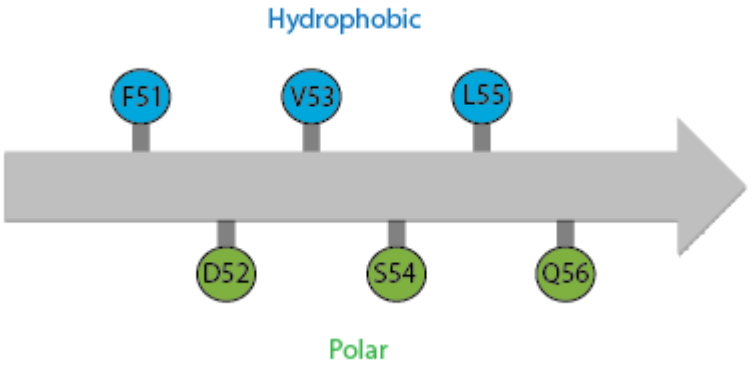


Figure 27: The *tapA* minimal functional unit encodes a β-strand from amino acids 51-56. Analysis of the TapA protein sequence using the secondary structure prediction software, Jpred 4, led to the discovery of a potential β-strand region constituted by amino acid residues 51-56. A schematic of the predicted β-strand structure 51-FDVSLQ-56 demonstrating the hydrophobic (blue circles) and polar (green circles) faces created by the side chains of the amino acids is shown.

Site-directed mutagenesis was used to generate *tapA*$_{1-57}$ ORFs encoding variants with amino acid substitutions at the desired positions. To elucidate the function of both conserved amino acids and the β-strand region found within TapA$_{1-57}$. A summary of the full-set of amino acid substitutions carried out is shown in **Figure 28**A. As before, the *tapA* variants generated were introduced to the *amyE* locus in the Δ*tapA* mutant background. The strains generated were

then grown under biofilm promoting conditions for 48 h to induce the expression of the *tapA* variants. For the purposes of comparison a number of control strains were grown and the biofilms formed as previously described (**Figure 28**B, C, D, E and F). The Δ*tapA* + *pIPTG+tapA*$_{1\text{-}57(F45A)}$ strain formed flat, featureless biofilms which resembled the Δ*tapA* mutant (**Figure 28**G). Biofilms formed by Δ*tapA* + *pIPTG+tapA*$_{1\text{-}57(D47A)}$ cells displayed partial recovery of structure (**Figure 28**H). In short, the amino acid substitutions of these 2 conserved residues perturbed the function of TapA, with a more profound effect observed when F45 was substituted with alanine.

The residues F51, D52, V53, S54, L55, Q56 and T57 are predicted to form a β-strand in which alternating residues point in the same direction to form a hydrophobic face (F, V and L) and a polar/charged face (D, S and Q) (**Figure 27**). Initially, an alanine scan was carried out targeting these residues and subsequent substitutions were assessed for selected amino acids. The strains Δ*tapA* + *pIPTG+tapA*$_{1\text{-}57(F51A)}$ (**Figure 28**I) and Δ*tapA* + *pIPTG+tapA*$_{1\text{-}57(D52A)}$ (**Figure 28**J) formed biofilms with only partial structuring. As D52 is a negatively charged polar amino acid it was judged that substituting this residue with the uncharged, polar asparagine would indicate whether the charge of the residue was important for function. The strain Δ*tapA* + *pIPTG+tapA*$_{1\text{-}57(D52N)}$ formed biofilms with partial structure with some level of wrinkling re-instated (**Figure 28**K). This is consistent with the previous result where Δ*tapA* + *pIPTG+tapA*$_{1\text{-}57(D52A)}$ did not display observable wrinkles. Taken together this indicated that the removal of the side chain (D52A) has more impact than removing the charge but retaining polar properties (D52N). In the cloning process a D52L mutant arose by chance and so it was carried forward for phenotyping. Biofilms of Δ*tapA* + *pIPTG+tapA*$_{1\text{-}57(D52L)}$ had only partial structure which demonstrates

that substitution of a polar amino acid with a hydrophobic amino acid is detrimental to TapA activity (**Figure 28**L). The Δ*tapA* + *pIPTG+tapA*$_{1\text{-}57(V53A)}$ biofilms showed partial structuring (**Figure 28**M). However, changing V53 to K, which is a charged polar residue, resulted in biofilms resembling Δ*tapA* in morphology (Δ*tapA* + *pIPTG-tapA*$_{1\text{-}57(V53K)}$) (**Figure 28**N). These results implied that replacing hydrophobic valine with the charged and polar lysine was detrimental to the activity of the TapA protein encoded, even more so than removing the side chain (V53A). Replacing valine with an alternative hydrophobic amino acid in isoleucine results in biofilms with wild-type morphology as displayed by the Δ*tapA* + *pIPTG-tapA*$_{1\text{-}57(V53I)}$ strain (**Figure 28**O). These results indicated that the hydrophobic side chain of valine is important for the function of TapA$_{1\text{-}57}$. Biofilms formed by Δ*tapA* + *pIPTG- tapA*$_{1\text{-}57(S54A)}$ resembled wild-type biofilms (**Figure 28**P). This result indicated that the polar nature of the serine side chain is not important for the function of TapA$_{1\text{-}57}$, however, like serine, alanine has a small R group the size of which may be important for function. Leucine at position 55 was changed to alanine and lysine, in both cases this resulted in the loss of TapA biofilm activity as demonstrated by the flat, featureless biofilms formed by Δ*tapA* + *pIPTG+tapA*$_{1\text{-}57(L55A)}$ (**Figure 28**Q) and Δ*tapA* + *pIPTG+tapA*$_{1\text{-}57(L55K)}$ (**Figure 28**R). However, substitution of leucine with the similarly hydrophobic isoleucine resulted in a functional variant of TapA$_{1\text{-}57}$ as displayed by the restoration of complexity in Δ*tapA* + *pIPTG-tapA*$_{1\text{-}57(L55I)}$ biofilms (**Figure 28**S). This result suggested that it is the hydrophobic nature of leucine which is important for the function of TapA$_{1\text{-}57}$. Changing T57 and Q56 to alanine, separately, did not impact TapA activity in restoring biofilm formation, as shown by Δ*tapA* + *pIPTG+tapA*$_{1\text{-}57(Q56A)}$ (**Figure 28**T) and Δ*tapA* + *pIPTG+tapA*$_{1\text{-}57(T57A)}$ (**Figure 28**U) biofilm phenotypes. This

indicated that it is likely to be the length of the protein that is important for function and not the biochemical properties of these 2 residues since $TapA_{1-56}$ is only partially functional (**Figure 28**F). A summary of the qualitative descriptions of the biofilm phenotypes for this experiment are shown in **Table 13**.

In summary, changing T57 or Q56 to alanine did not impact biofilm formation indicating that it is the length of the protein that is important for function and not the biochemical properties of these 2 residues given that $TapA_{1-56}$ is not functional. The hydrophobic character of V53 and L55 was shown to be important for the function of $TapA_{1-57}$. The residues F51 and D52 were shown to be key residues as changing these residues perturbed TapA activity. S54 appeared to be the most dispensible of the amino acids analysed as removing the side chain properties by replacement with alanine (S54A) did not perturb the biofilm acitivity of TapA.

It should be noted that the antibody cannot detect truncated forms of TapA and, therefore, we cannot distinguish a lack of protein production from a loss of function. However, these results indicate support for the hypothesis that the hydrophobic face of a potential β-strand structure within the TapA minimal functional unit is important for activity.

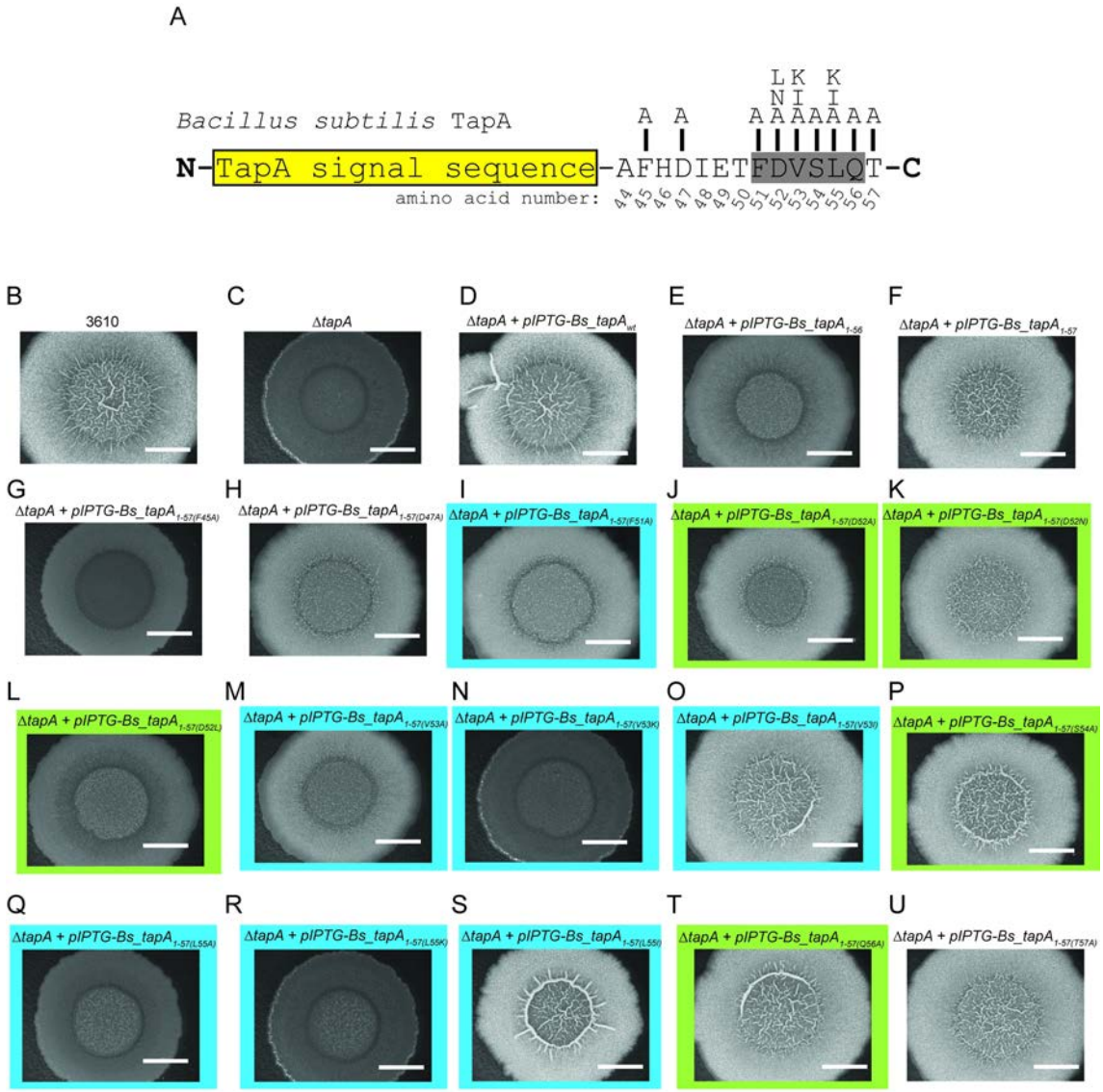


Figure 28: Mutational analysis of the *tapA* minimal functional unit. (A) A summary of the amino acid substitutions of the $TapA_{1-57}$ which are carried out as part of this experiment. Shown is the *B. subtilis* TapA sequence and in yellow is the predicted signal peptide. Highlighted in grey is the β-strand region. The letters shown above the sequence are the amino acids which were substituted in place of those found in the TapA sequence. The amino acid numbering is shown below and the N- and C-termini are labelled in bold. Biofilms formed by a range of strains are shown in: **(B)** 3610, **(C)** Δ*tapA* **(D)** Δ*tapA* + *pIPTG-tapA*$_{wt}$ **(E)** Δ*tapA* + *pIPTG-tapA*$_{1-56}$ (NRS6025) **(F)** Δ*tapA* + *pIPTG-tapA*$_{1-57}$ (NRS6041) **(G)** Δ*tapA* + *pIPTG+tapA*$_{1-57(F45A)}$ **(H)** Δ*tapA* + *pIPTG+tapA*$_{1-57(D47A)}$ **(I)** Δ*tapA* + *pIPTG+tapA*$_{1-57(F51A)}$ **(J)** Δ*tapA* + *pIPTG+tapA*$_{1-57(D52A)}$ **(K)** Δ*tapA* + *pIPTG+tapA*$_{1-57(D52N)}$ **(L)** Δ*tapA* + *pIPTG+tapA*$_{1-57(D52L)}$ **(M)** Δ*tapA* + *pIPTG+tapA*$_{1-57(V53A)}$ **(N)** Δ*tapA* + *pIPTG-tapA*$_{1-57(V53K)}$ **(O)** Δ*tapA* + *pIPTG-tapA*$_{1-57(V53I)}$ **(P)** Δ*tapA* + *pIPTG- tapA*$_{1-57(S54A)}$ **(Q)** Δ*tapA* + *pIPTG+tapA*$_{1-57(L55A)}$ **(R)** Δ*tapA* + *pIPTG+tapA*$_{1-57(L55K)}$ **(S)** Δ*tapA* + *pIPTG-tapA*$_{1-57(L55I)}$ **(T)** Δ*tapA* + *pIPTG+tapA*$_{1-57(Q56A)}$ **(U)** Δ*tapA* + *pIPTG+tapA*$_{1-57(T57A)}$. Blue boxes signify amino acid positions which are hydrophobic in the WT protein sequence and green boxes indicate amino acid positions which are polar in the WT sequence. See Table **13** for full-details of the strains used and the corresponding strain numbers. Strains carrying inducible

tapA variants were grown in the presence of 25 µM IPTG. *Bs_tapA* = *B. subtilis tapA*. Scale bar = 0.5 cm. For figure **(B)**-**(U)** n=3.

***tapA* variant**[a]	**Biofilm phenotype**[b]	**Strain number**
Bs_tapA$_{1\text{-}56}$	Partial/defect	NRS6025
Bs_tapA$_{1\text{-}57}$	WT	NRS6041
Bs_tapA$_{1\text{-}57(T57A)}$	WT	NRS6384
Bs_tapA$_{1\text{-}57(Q56A)}$	WT	NRS6385
Bs_tapA$_{1\text{-}57(L55I)}$	WT	NRS6386
Bs_tapA$_{1\text{-}57(L55K)}$	Defect (Δ*tapA* phenotype)	NRS6387
Bs_tapA$_{1\text{-}57(L55A)}$	Defect (Δ*tapA* phenotype)	NRS6472
Bs_tapA$_{1\text{-}57(S54A)}$	WT	NRS6388
Bs_tapA$_{1\text{-}57(V53I)}$	WT	NRS6389
Bs_tapA$_{1\text{-}57(V53K)}$	Defect (Δ*tapA* phenotype)	NRS6502
Bs_tapA$_{1\text{-}57(V53A)}$	Partial/defect (like Bs_TapA$_{1\text{-}56}$)	NRS6473
Bs_tapA$_{1\text{-}57(D52A)}$	Partial	NRS6476
Bs_tapA$_{1\text{-}57(D52L)}$	Partial/defect (like Bs_TapA$_{1\text{-}56}$)	NRS6477
Bs_tapA$_{1\text{-}57(D52N)}$	Partial (wrinkling can be observed)	NRS6516
Bs_tapA$_{1\text{-}57(F51A)}$	Partial	NRS6390
Bs_tapA$_{1\text{-}57(D47A)}$	Partial	NRS6475
Bs_tapA$_{1\text{-}57(F45A)}$	Defect (Δ*tapA* phenotype)	NRS6474

Table 13: Biofilm phenotypes with qualitative description of the results of mutational analysis of the B. subtilis *tapA* minimal functional genetic unit. [a] Origin strain of the *tapA* sequence used is denoted as follows: *Bs_tapA* = *tapA* from *B. subtilis*. [b] Biofilm phenotypes are described qualitatively as: WT- prominent corrugations resembling WT *B. subtilis* biofilms, partial– failure to form prominent wrinkles and defect– flat, featureless biofilms resembling a Δ*tapA* mutant biofilm phenotype.

3.3.7. The *tapA* variants encoding non-functional TapA$_{1\text{-}57}$ variants do not appear to display a dominant negative effect

It was then investigated whether the *tapA* variants judged to encode non-functional products (*tapA*$_{1\text{-}57(F45A)}$, *tapA*$_{1\text{-}57(L55K)}$ and *tapA*$_{1\text{-}57(L55A)}$) had a dominant negative effect on biofilm formation when expressed from the *amyE* locus in wild-type 3610 biofilms. Biofilm formation was observed as described previously. The wild-type 3610 biofilm was included as a positive control and displayed complex, rugose morphology and the Δ*tapA* mutant exhibited a flat, featureless appearance (**Figure 29**). The strains 3610 + pIPTG- *tapA*$_{1\text{-}57(F45A)}$, 3610 + pIPTG- *tapA*$_{1\text{-}57(L55K)}$ and 3610 + pIPTG- *tapA*$_{1\text{-}57(L55A)}$ were grown on

MSgg agar containing IPTG to induce expression of the *tapA* variants (**Figure 29**). All of the biofilms displayed wild-type morphology. This result does not support the hypothesis that these forms of TapA exert an inhibitory effect on the native TapA protein. However, the presence of these truncated TapA variants *in vivo* has not been proven and they may be degraded by the cells during biofilm formation.

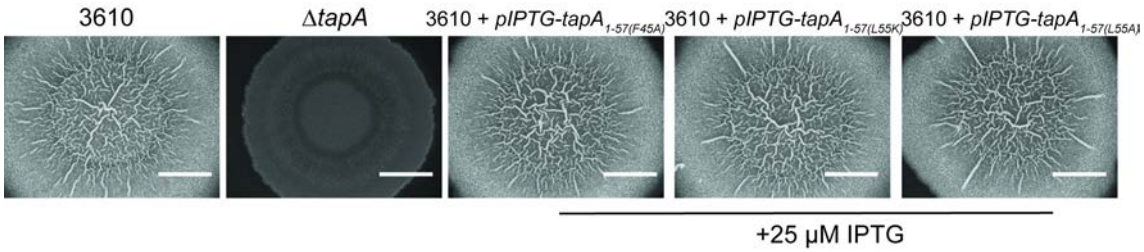


Figure 29: Biofilm morphology of 3610 strains expressing versions of the *tapA* gene encoding variants of the minimal functional unit. Expression of *tapA* variants was induced in the presence of 25 µM IPTG. The following strains were used: 3610, Δ*tapA* (NRS3936), 3610 + *pIPTG-tapA*$_{1\text{-}57(F45A)}$ (NRS6522), 3610 + *pIPTG-tapA*$_{1\text{-}57(L55K)}$ (NRS6520) and 3610 + *pIPTG-tapA*$_{1\text{-}57(L55A)}$ (NRS6521). Scale bar = 0.5 cm. n=1.

3.3.8. Determining the functional form of TapA *in vivo*

It has been shown above that TapA$_{1\text{-}57}$ is functional, additionally, TapA is predicted to have a cleavable signal peptide consisting of amino acids 1-43, this suggested that TapA may function as a short, secreted peptide consisting of amino acids 44-57 (**Figure 30**). To test this hypothesis 2 approaches were taken. Firstly, the ability of the TapA$_{44\text{-}57}$ peptide supplied exogeneously to restore wild-type morphology to Δ*tapA* mutants was tested. Secondly, a genetic construct encoding a hybrid protein consisting of the known signal peptide of TasA fused to the TapA$_{44\text{-}57}$ peptide region was constructed and tested for biofilm activity.

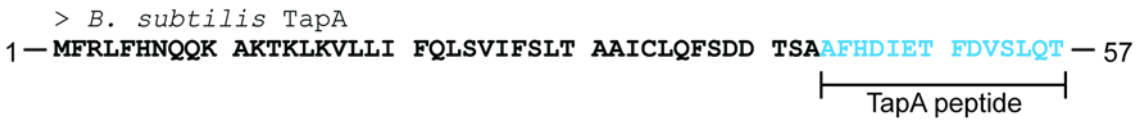


Figure 30: A schematic of a potential TapA peptide. Shown is the amino acid sequence of $TapA_{1-57}$ encoded by the *tapA* minimal functional unit. Bold and underlined is the TapA signal sequence which is proposed to be cleaved. In light blue is $TapA_{44-57}$ which would theoretically be released upon cleavage of the signal sequence.

3.3.9. The $TapA_{44-57}$ peptide does not demonstrate activity when added exogeneously to Δ*tapA* biofilms

The $TapA_{44-57}$ peptide was synthesised and the ability to restore structure to Δ*tapA* (NRS3936) biofilms was determined. Approximately 80 µg of the $TapA_{44-57}$ peptide was added to the culture before spotting of the cells onto biofilm promoting MSgg media. 3610 biofilms formed complex, structured biofilms on the provision of purification buffer (+buffer) or $TapA_{44-57}$ (+peptide) (**Figure 31**A). As expected the Δ*tapA* + *pIPTG-tapA* (NRS5045) strain biofilms resembled wild-type on induction of *tapA* expression with IPTG (**Figure 31**A). In short the biofilms formed as normal regardless of the provision of buffer or peptide. In the case of Δ*tapA* biofilms, the provision of purification buffer or exogeneous peptide had no impact on the gross biofilm features (**Figure 31**A).

There are many reasons why exogeneous provision may not rescue biofilm formation. For example, since the peptide is being added exogeneously the correct localisation may not be achieved by comparison with the endogeneously produced TapA. Therefore a second approach was taken to examine whether TapA functions as a peptide.

3.3.10. The hybrid protein $TasA_{sp}$-$TapA_{44-57}$ does not display biological activity in the biofilm

To test whether a cleavable signal peptide was needed for the export and function of the TapA protein, the following experiment was carried out. The known SipW-cleaved signal peptide of TasA ($TasA_{1-27}$) was fused to the region of *tapA* encoding amino acids 44-57 (**Figure 31**B). When this artificial construct, encoding $TasA_{1-27}TapA_{44-57}$, was analysed *in silico* by SignalP4.1 it resulted in a predicted cleavage point, directly after amino acid 27, with a higher score (D= 0.644), than that obtained for WT TapA (D = 0.527). When the construct encoding this protein was induced, in the strain Δ*tapA* + *pIPTG-tasA*$_{1-27}$*tapA*$_{44-57}$, no activity was observed with the resulting colonies resembling the flat Δ*tapA* mutant biofilms (**Figure 31**C).

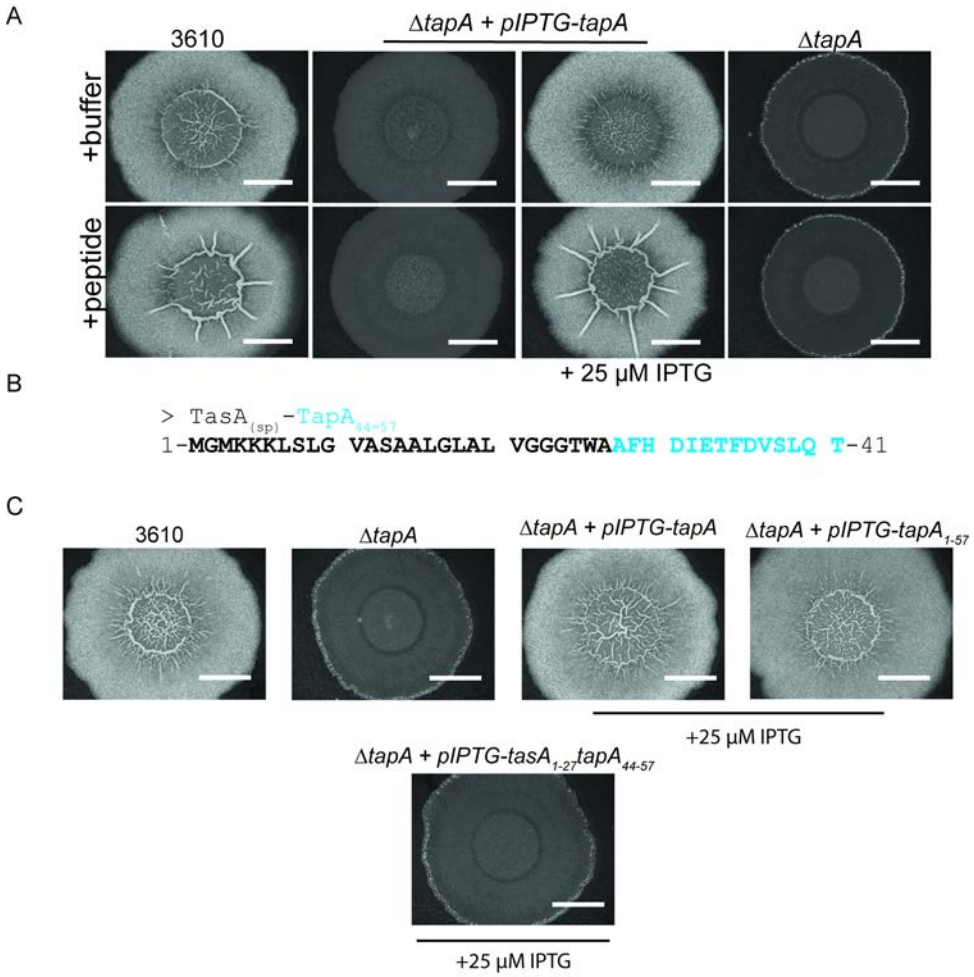


Figure 31: The peptide TapA$_{44-57}$ does not restore complexity to Δ*tapA* biofilms. (A) Bacterial cultures were grown for 3-4 h before cells were collected by centrifugation to concentrate and then mixed 1:1 (vol:vol) with purification buffer (+buffer) or with peptide dissolved in purification buffer (+ peptide). Biofilms were spotted onto MSgg agar with or without the presence of 25 μM IPTG and grown for 48 h at 30^{o}C. Strains 3610, Δ*tapA* (NRS3936) and Δ*tapA* + *pIPTG-tapA* (NRS5045) were used as controls. n=1. **(B)** A schematic outlining the protein encoded by the *tasA$_{1-27}$tapA$_{44-57}$* construct. Shown in bold, black font is the TasA signal peptide. In blue is the region of TapA from amino acids 44-57. **(C)** Biofilm morphology of Δ*tapA* strains expressing IPTG-inducible constructs encoding *tasA-tapA* hybrid sequences, imaged at 48 h. The TasA signal sequence was fused to the N-terminus of the predicted mature (post-signal peptide cleavage) form of TapA from amino acids 44-57. n=2. Strains: 3610, Δ*tapA* (NRS3936) *ΔtapA* + *pIPTG-tapA* (NRS5045)*, ΔtapA* + *pIPTG-tapA$_{1-57}$* (NRS6041) and *ΔtapA* + *pIPTG-tasA$_{1-27}$tapA$_{44-57}$* (NRS6478). Scale bar = 0.5 cm.

Controls demonstrated that the IPTG-dependent induction of gene expression was functioning in the assay with both Δ*tapA* + *tapA* (NRS5045) and Δ*tapA* + *tapA$_{1-57}$* (NRS6041) forming biofilms resembling wild-type in appearance. These

results cast doubt on whether the N-terminus of TapA forms a cleavable signal peptide.

3.3.11. The extracellular processing of TapA is not dependent on SipW-mediated export

A cleavable signal sequence from amino acids 1-43 is predicted *in silico* for TapA. SipW has been reported to be the cognate signal peptidase that carries out cleavage of both TapA and TasA signal peptides (Stöver and Driks, 1999a, c). However, attempts to demonstrate the functionality of the peptide $TapA_{44-57}$ remaining after the anticipated cleavage of the $TapA_{1-57}$ have so far been unsuccessful. A possible explanation for this could be that amino acids 1-43 are not removed from, but instead part of the functional TapA form *in vivo*. To address this possibility 2 approaches were taken, first the *sipW* gene was deleted and the impact of this on biofilm formation and TapA export examined. Second, a form of TapA was constructed in which the cleavage site is changed to block the action of the *B. subtilis* signal peptidases (of which there are 5). The second approach is based on the report that proline residues are not tolerated in the +1 position, in relation to the cleavage site, targeted by signal peptidase enzymes (Barkocy-Gallagher and Bassford, 1992). For this reason a *tapA* variant was constructed to encode the full-length TapA protein with a proline residue at amino acid position 44, in place of, the native alanine ($tapA_{A44P}$). Biofilms were grown as described previously. The biofilms formed by the Δ*sipW* (NRS5488) strain resembled the flat, featureless appearance of Δ*tapA* biofilms, in contrast to the rugose, complex colonies formed by the parental strain 3610 (**Figure 32**A). As expected, induction of the wild-type *tapA* coding region in the strain Δ*tapA* + *pIPTG-tapA* (NRS5045) (**Figure 32**B) restored biofilm formation to that displayed by the 3610 strain (**Figure 32**A). Interestingly, when the Δ*tapA*

+ *pIPTG-tapA*$_{A44P}$ (NRS6059) strain was grown in the presence of IPTG, biofilm architecture was also re-instated to wild-type levels (**Figure 32**B). By contrast, in the absence of IPTG, Δ*tapA* + *pIPTG-tapA*$_{A44P}$ (NRS6059) biofilms were flat and featureless (**Figure 32**B). These findings reveal that a change of alanine 44 to proline does not impact TapA activity.

Next it was tested whether TapA processing is perturbed in either the Δ*sipW* or Δ*tapA* + *pIPTG-tapA*$_{A44P}$ (NRS6059) strains. To test this, the biofilm lysates of these strains were analysed and probed with the αTapA antibody. As shown previously 3 bands were detected by the αTapA antibody in the 3610 biofilm lysates, all of which were absent in the Δ*tapA* sample (**Figure 32**C). The 3 bands were restored on induction of *tapA* in Δ*tapA* + *pIPTG-tapA* (NRS5045) (**Figure 32**C). In the Δ*sipW* lysates only the bottom 2 bands were detected, these correspond to the low molecular weight form of TapA (16 kDa) and the TasA/TapA-dependent band (18 kDa) (**Figure 32**C). All three bands were present in the Δ*tapA* + *pIPTG-tapA*$_{A44P}$ lysate originating from biofilms grown with IPTG present (**Figure 32**C). The 3 bands were also found in the Δ*tapA* + *pIPTG-tapA*$_{A44P}$ grown without IPTG (**Figure 32**C). The IPTG-inducible P_{spank} promoter used in this study is known to drive residual levels of expression in the absence of IPTG which explains this finding. In both cases these results indicated that TapA still achieved export to an extracellular location in order to be processed to low molecular weight forms by the action of exoproteases, as proposed in section **3.1.16**. Additionally, it was interesting to note that the non-specific TasA band normally detected by the αTapA antibody was no longer observed in the Δ*sipW* lysates. This is likely to be due to the role of SipW in regulating the export of TasA (Stöver and Driks, 1999c). To validate these findings an αTasA immunoblot was carried out on 3610, Δ*tapA* (NRS3936),

Δ*tasA* (NRS5267) and Δ*sipW* (NRS5488) biofilm lysates (**Figure 32**D). TasA was detected as a strongly reacting band of ~30 kDa in 3610 lysates, with a number of lower molecular weight forms found below (**Figure 32**D). As expected, TasA levels were reduced in the Δ*tapA* mutant and absent in the Δ*tasA* lysates. However, only a limited amount of TasA could be detected in Δ*sipW* biofilms (**Figure 32**D). It may be that TasA could not be exported without SipW and was then subjected to intracellular degradation. These results are consistent with a role for SipW in processing the signal peptide of TasA (Stöver and Driks, 1999c).

To summarise, SipW may not be required for the export of TapA to an extracellular location. Furthermore, attempts to block signal peptide cleavage of TapA by any of the signal peptidases encoded by *B. subtilis* did not impact biofilm formation or TapA processing. In conclusion, it still remains unclear whether TapA functions as a peptide ($TapA_{44-57}$) or whether amino acids 1-43 are a present in the mature, functional form of TapA.

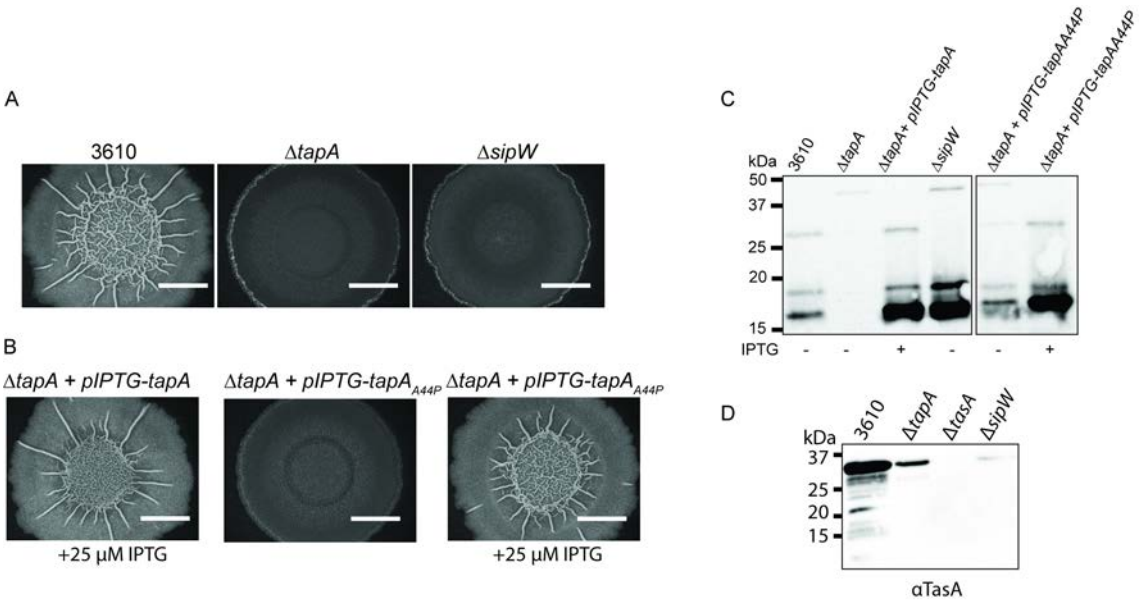


Figure 32: TapA localization does not appear to be dependent on an N-temrinal signal peptide. (A) Biofilm formation was assessed in the following strains: 3610, Δ*tapA* (NRS3936), Δ*sipW* (NRS5488) (n=>3), **(B)** Δ*tapA* + *pIPTG-tapA* (NRS5045) and Δ*tapA* + *pIPTG-tapA*$_{A44P}$

(NRS6059). n=1. **(C)** Biofilm lysates (50 µg) of strains shown in (A) and (B) were separated on a 14% acrylamide gel by SDS-PAGE, transferred to a membrane and probed with the αTapA antibody for immunoblot. n=1. **(D)** αTasA immunoblot of biofilm samples: 3610, Δ*tapA*, Δ*tasA* and Δ*sipW* biofilm lysate. n=2. Scale bar = 0.5 cm. - = 0 µM IPTG + = 25 µM IPTG.

3.3.12. Summary

This results chapter has outlined a genetic approach which was undertaken to understand the function of TapA. It was found that homologous *tapA* genes originating from *B. pumilis, B amyloliquefaciens* and *B. paralicheniformis* encoded functional products when studied in the context of *B. subtilis* 3610. Strikingly, a very well conserved region of the protein from amino acids 111-120 was not needed for the biofilm activity of TapA. Furthermore, it was identified that most of the C-terminal portion of the protein was not needed for the function of TapA. On further analysis, it was found that a minimal functional unit of *tapA* encoding just the first 57 amino acids of TapA ($TapA_{1-57}$) was sufficient to restore biofilm formation to a Δ*tapA* mutant.

A site-directed mutagenesis approach highlighted amino acid residues of importance present in $TapA_{1-57}$, and a potential β-strand region of interest was investigated further. Given the prediction of an N-terminal signal peptide from amino acids 1-43 this indicated that TapA may function as a peptide consisting of amino acids 44-57. A functional study of the TapA peptide did not confirm activity and a re-assessment of the TapA signal peptide indicated that it may instead form a part of the mature, functional form of TapA.

4. Discussion

Each component of the matrix operon (*tapA-sipW-tasA*) has been shown to be essential for complex biofilm formation in *B. subtilis* (Branda et al., 2006; Hamon et al., 2004; Romero et al., 2011; Terra et al., 2012). The current work focussed on the secreted matrix protein TapA, which is reported to facilitate fibre formation by TasA and to anchor these fibres onto the surface of the cell. We wanted to better understand the function of TapA and how it interacts with TasA to facilitate fibre formation. Additionally, we wanted to determine the composition of the functional form of TapA *in vivo*.

Using a *de novo* αTapA antibody we found that TapA is processed to a low molecular weight form in the biofilm (**Figure 13**). Furthermore, we found that processing is dependent on the action of the natively encoded exoproteases (**Figure 20**). We also demonstrated that the presence of the *tapA* gene is not necessary for the biological function of TasA fibres formed *in vitro* (**Figure 22**). Next, we identified that the minimal functional form of *tapA* encoded the first 57 amino acids of TapA ($TapA_{1-57}$) (**Figure 24**). This form of the protein is shown in relation to the predicted 3D structure of TapA (**Figure 33**A) and schematically (**Figure 33**C) in relation to full-length TapA (**Figure 33**B and **Figure 33**C). It was found that the αTapA antibody did not recognise the minimally functional form of the TapA protein (**Figure 25**). Therefore, it is likely that the low molecular weight form of TapA, detected by immunoblot, is similar to the core of the protein (predicted to be a structured β-strand rich region **Figure 33**A) which is absent in the minimal functional form (**Figure 33**C). In support of this conclusion, it was found that when a well conserved motif between amino acids 111-120, which resides within this core region, was changed, then the low molecular weight forms of TapA could no longer be detected (**Figure 23**E). The core of the protein was shown to be resistant to proteolysis by exoproteases

found in supernatant and stable over time when $TapA_{44-253}$ was stored at 4°C. These forms are shown schematically (**Figure 33**D and E), note that they largely overlap with the region thought to contain the β-strand rich domain. The stability of this part of the protein may explain why it was readily detectable *in vivo*. We then went on to highlight a predicted β-strand structure, within the minimally functional region, which appears to be key for the function of $TapA_{1-57}$ (**Figure 28**), this is shown in the context of the predicted tertiary structure for TapA (**Figure 33**A). Finally, we re-assessed the role of the signal peptidase, SipW, in the processing of TapA and re-investigated the existence of an N-terminal TapA signal peptide.

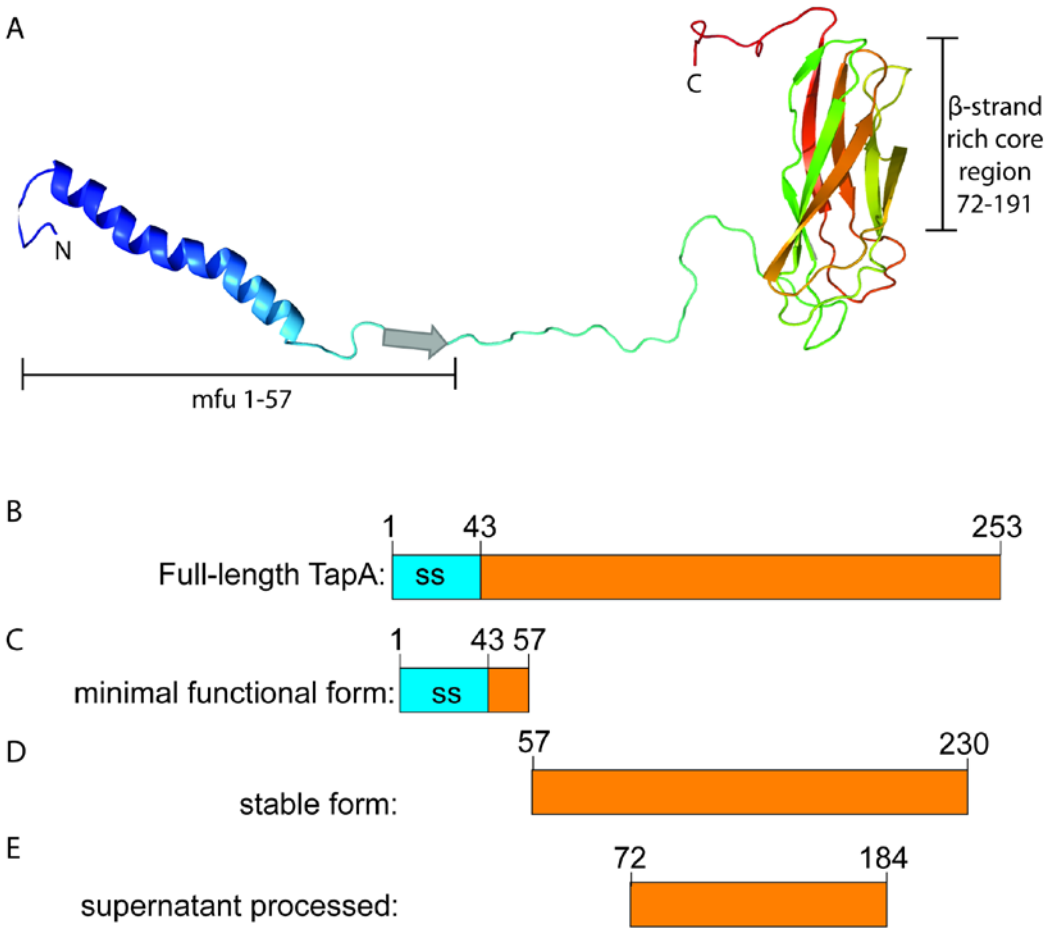


Figure 33: The 3D structure of TapA annotated with functionally and structurally important features (A) The 3D structure of TapA as predicted using the RaptorX server and visualised using PyMol. The N- and C-termini are labelled. The form of the minimal functional

unit (mfu) is shown on this structure with the predicted β-strand secondary structure element shown as a grey arrow. The structured β-strand rich domain is also shown. Schematics illustrate different forms of the TapA protein important in the current study. The full-length protein **(B)**, minimal functional form **(C)**, the stable form which persists on storage at 4^{o}C **(D)** and the supernatant processed form **(E)** are shown. In all cases numbers are in relation to the amino acid numbering of the WT, full-length protein. In the schematic representations, the signal sequence (ss) is shown in blue and the mature region of TapA is shown in orange.

4.1.1. Re-evaluating the functional form of TapA

Using a custom made αTapA antibody it was found that TapA is processed in the biofilm environment, migrating as a low molecular weight band of 16 kDa, when analysed by immunoblot (**Figure 13**B). This is in contrast to the higher molecular weight TapA bands detected by other groups (Romero et al., 2011, 2014; Stöver and Driks, 1999a). It was also demonstrated that the development of an αTapA antibody is made problematic due to cross-reactivity of the antibody to the more abundant protein TasA. This result was consistent with the result observed when using the earlier published TapA anti-serum which was kindly gifted by Adam Driks at Loyola University (**Figure 13**C). Interestingly, when concentrated TasA fibres were probed with an αTapA antibody, 2 smaller bands migrating between the 15 kDa and 25 kDa markers are detected, which are reminiscent of the bands I observe (Romero et al., 2011) (Figure S5). Furthermore, the nature of the samples probed with αTapA antibodies vary between studies for example the current study utilises biofilms that are exposed to an enzymatic lysis buffer and sonication and contain intracellular proteins, membrane proteins and extracellular proteins. However, other studies utilize concentrated cell fractions to give localisation information for TapA.

It was then shown that TapA is processed in an active manner which is dependent upon the activity of secreted proteases found in the biofilm (**Figure 20**B). Generation of an exoprotease-free version of NCIB3610 *comI*, which did

not display exoprotease activity, as judged by growth on LB agar supplemented with milk (**Figure 20**A), resulted in the detection of unprocessed TapA (**Figure 20**B). However, there was no biofilm defect observed in this strain (**Figure 20**C). This was a suprising result given that exoprotease genes are known to be expressed within the biofilm (Marlow et al., 2014) and were shown to process TapA (**Figure 20**B). These results, however, indicate that the role of exoproteases may not be in facilitating the formation of architecturally complex colonies under these conditions. The possibility does remain that there is sufficient protease activity within the biofilm to generate a very small amount of a processed, functional form of TapA. This hypothesis is supported by the finding that the supernatant of 3610 *comI*-KO7 strain showed residual levels of $TapA_{44\text{-}253}$ proteolytic activity (**Figure 20**D). An alternative role for exoproteases may be the breakdown of biopolymers to release energy for the maintenance of cellular processes and division. Our work does indicate that the low molecular weight form of TapA is dependent on the action of the exoprotease gene *vpr*. It is possible that low molecular weight TapA is generated by the proteolytic cleavage of the 18 kDa band, detected by the αTapA antibody. Therefore, the 16 kDa and 18 kDa bands are both TapA, with the latter being dependent on the presence of TasA. Regardless of the identity of the 18 kDa band it appears that the presence of the 16 kDa band is dependent on the action of the exoprotease Vpr (**Figure 20**B and **Figure 18**B). However, this gene should be deleted independently to generate a Δ*vpr* single mutant to test this hypothesis further. This is because in the current work the *vpr* gene is deleted within the context of other exoprotease genes being deleted at the same time, so a combinatorial role between Vpr and another exoprotease such as Bpr, to generate TapA at 16 kDa, cannot be discounted.

4.1.2. An alternative approach to studying TapA validates amino acids 50-57 as a key region for TapA function

Findings from this project led to the conclusion that amino acids 51-57 of TapA are representative of a region which is crucial for bioactivity in the NCIB3610 biofilm, as previously published (Romero et al., 2014). Importantly, this finding was elucidated using an independent approach. Previously, a functional role for this region was predicted based on the observation that it represented an imperfect repeat sequence which was found to be similar to a region at positions 194 to 237 in the protein. When amino acids 50-57 were deleted or replaced with a different set of amino acids then the protein was no longer functional (Romero et al., 2014). The approach taken in this study was to compare the protein sequence of *B. subtilis* TapA to orthologous TapA proteins from related *Bacillus* species to identify key sequence features. It was found that the expression of homologous *tapA* genes from *B. pumilis*, *B. amyloliquefaciens*, and *B. paralicheniformis* restored wild-type biofilm formation to a Δ*tapA* strain (**Figure 23**A). This result is in agreement with reports that *B. amyloliquefaciens tapA* can restore architecture to a *B. subtilis* Δ*tapA* mutant (Romero et al., 2014). It was noted that *Bacillus pumilis* TapA lacked a large part of the C-terminus but was still functional. Given that TapA was found to be processed *in vivo* and that parts of the sequence were not needed for function we looked to identify the minimal functional unit of *tapA*. We designed a range of *tapA* gene variants which encoded truncated TapA which lacked an increasing number of amino acids from the C-terminus. The result of this was the identification of the minimal functional genetic unit (mfu) determined to encode amino acids 1-57 (**Figure 24**A). This is in agreement with the finding

that amino acids 50-57 are key for the function of TapA (Romero et al., 2014). However, here we extend this finding to show that a region consisting of TapA from amino acids 1-57 is necessary and sufficient for activity within the biofilm.

4.1.3. Determining the minimal functional unit of TapA orthologues

We suggested that determining the minimal functional forms of the TapA orthologues could give insight into the essential, conserved features of TapA. The *B. pumilis*, *B. amyloliquefaciens*, and *B. paralicheniformis* TapA homologs equivalent in length to *B. subtilis* $TapA_{1-57}$ all had predicted β-strands within this region (**Figure 26**F). Constructs were also designed which encoded additional amino acids on the C-terminal side of what is equivalent to the $Bsub_{1-57}$ mfu. This is because it was not known whether the mfu for the *tapA* homologs is greater in length than the *B. subtilis* version. It was not possible to determine a minimal functional unit for *B. paralicheniformis* (**Figure 26**E) or *B. pumilis* (**Figure 26**D) based on the experiments conducted here. The *B. paralicheniformis* short forms of TapA did appear to have partial activity but did not restore wild-type biofilm formation (**Figure 26**E). However, *B. amyloliquefaciens* had a minimal functional form similar in length to that of *B. subtilis* TapA (**Figure 26**C). It should be noted that these genes are being expressed in a *B. subtilis* Δ*tapA* mutant which is not the system in which they have evolved. This may be why the *B. amyloliquefaciens* TapA truncations displayed more biological activity, as the sequence identity is highest between *B. amyloliquefaciens* and *B. subtilis* TapA. In support of this, we have published that recombinant *B. amyloliquefaciens* fTasA, added exogeneously to Δ*tasA sinR* (NRS5248) biofilms restored rugosity (Erskine et al., 2018). Interestingly, the *B. amyloliquefaciens* fibres were found to form a distributed mesh of thin

fibres, whereas, *B. subtilis* TasA fibres assembled in to bundles composed of fibres with a greater diameter (Erskine et al., 2018).

4.1.4. Mutational analysis of TapA 51-57 and conserved residues

The TapA amino acid sequence was analysed using the secondary structure prediction programme Jpred 4 (Drozdetskiy et al., 2015). This highlighted a possible 6 residue β-strand composed of amino acids 51-56 inclusive. It was hypothesised that the functionality of the $TapA_{1-57}$ variant is dependent on this secondary structure feature. If the prediction is accurate then it would result in alternating hydrophobic and polar residues, meaning that one face of the strand is hydrophobic and the opposite face polar. To test this, a range of mutant *tapA* forms were generated which encoded the first 57 amino acids of the protein with specific amino acid changes in the region of interest. Changing T57 to alanine and Q56 to alanine had no impact on biofilm integrity. This indicates that these residues play an important function in conveying length to the protein and that the biochemical properties of the residues are less important, this is due to the observation that $TapA_{1-56}$ has noticeably reduced function. However, it should be noted that threonine is similar to alanine in that it has a small and hydrophobic side chain. Carrying out an alanine scan, targeting the polar face, allowed us to identify the importance of residue D52 (**Figure 28**: J, K and L), however, the change of S54A and Q56A (as discussed above) had no impact on $TapA_{1-57}$ function (**Figure 28**: P and T). Changing the hydrophobic residues to alanine or to a positively charged lysine residue reduced or abolished $TapA_{1-57}$ activity (**Figure 28**: I, J, M, N, O, Q, R and S). This indicated that the hydrophobicity of these residues is important, however, the small, hydrophobic amino acid alanine cannot substitute in their place. These results are summarised in **Figure 34**. In conclusion, the hydrophobic properties of amino

acids 51, 53 and 55 are important. This hydrophobic face may be important for TasA interaction or for interaction with other matrix components. Additionally, conserved residues F45 and D47 were also shown to be needed for $TapA_{1-57}$ function (**Figure 28**: G and H). Aromatic amino acids such as phenylalanine (F45) have been shown to form stacking interactions which can be important for protein structure and function.

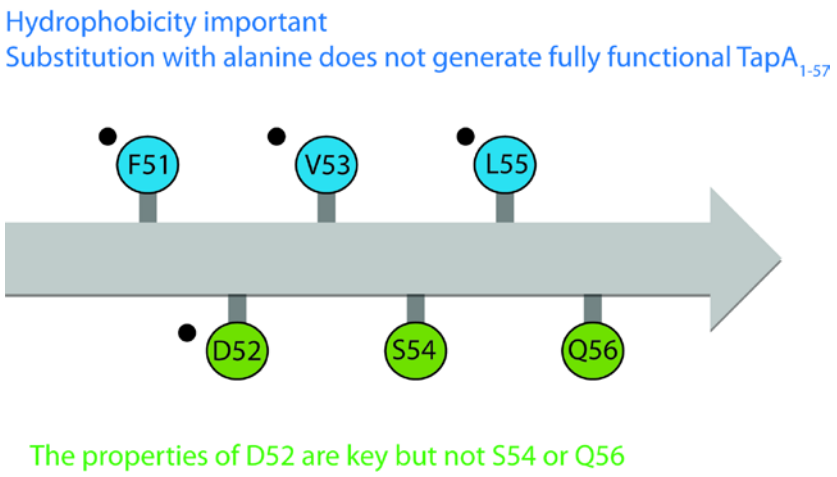


Figure 34: A model for the secondary structure for the region of TapA from amino acids 51-56 inclusive which are predicted to adopt a β-strand conformation (grey arrow). Hydrophobic residues are shown as blue circles and polar residues are shown as green circles. The amino acids which were shown to be key for the function of $TapA_{1-57}$ are indicated with a black circle.

It was reported that $TapA^{\Delta 50-57}$ had a dominant negative effect preventing normal biofilm formation (Romero et al., 2014). Additionally, this TapA variant was shown to be defective in assisting polymerisations of TasA fibres *in vitro* (Romero et al., 2014). It was therefore hypothesised that the non-functional $TapA_{1-57}$ variants generated in this project may have a similar effect. To test this *Bs_tapA*$_{1\text{-}57(L55K)}$, *Bs_tapA*$_{1\text{-}57(L55A)}$ and *Bs_tapA*$_{1\text{-}57(F45A)}$ were expressed from the *amyE* locus under the control of the *Pspank* promoter in the NCIB3610 strain background. These strains formed complex colonies resembling wild-type (**Figure 29**). This indicates that they encode products which do not have a

dominant negative effect. It should be noted that the TapA$^{\Delta 50\text{-}57}$ form was a full-length form of TapA, unlike the truncated forms analysed in this work.

Ideally, when carrying out the genetic analysis of the *tapA* sequence, which are designed to encode variants of the TapA protein, and testing them for function, then an αTapA immunoblot should be carried out. This is to determine that the protein is still being made and is found at similar levels to native TapA. However, as demonstrated, the αTapA antibody utilised does not recognise truncated forms of the protein. To remedy this, an antibody could be generated to recognise the minimal functional form of the protein. Alternatively, the translational fusion of TapA variants with an epitope tag may allow for detection with an alternative antibody. It has been shown that the C-terminus of full-length TapA could be tagged with yellow fluorescent protein (YFP) and detected in the cell wall fraction of biofilm cells (Romero et al., 2011). It is unclear how this was successful given the *in vivo* processing of TapA, which we have demonstrated. However, it could be that there is a small amount of full-length TapA that we can not detect directly with the αTapA antibody. Alternatively, the C-terminus of TapA may be cleaved from the functional N-terminus and interact with the cell wall. Attempting to fuse the full-length TapA, TapA$_{1\text{-}193}$, and TapA$_{1\text{-}57}$ to a C-terminal His-tag did not appear to be successful (data not shown). Although it is interesting to note that TapA$_{1\text{-}57_6His}$ was still functional given that there is an additional 6 amino acids added to the protein. It could indicate that the length of the protein is more important than the biochemical properties of the amino acids after threonine-57. Consistent with this, changing amino acids 56 and 57 to alanine did not impact function. Finally, an αTasA immunoblot could be carried out to elucidate whether TapA variants retain the property of facilitating TasA stability/production. The role of TapA in TasA stability was a phenomenon that

we observed (**Figure 11**E) and has been reported separately (Romero et al., 2011)

4.1.5. TapA is not needed for *ex vivo* TasA fibre function

The restoration of biofilm architecture to a Δ*tasA sinR* strain by *ex vivo* provision of *in vitro,* self-assembled TasA fibres was shown not to be dependent on the presence of the *tapA* gene (**Figure 22**B) (Erskine et al., 2018). This is in addition to our observation that exogenous provision of 20 µg of fTasA is unable to restore complex architecture to the *tasA* mutant (**Figure 21**B) (Erskine et al., 2018). In contradiction to this result, it has been reported that the addition of 40 µg of recombinant TasA fibres to Δ*tasA* pellicles can restore structure to a level indistinguishable from wild-type (Branda et al., 2006), and this result has been reproduced in a separate study (Diehl et al., 2018). Additionally, a preparation of native TasA was also shown to have biological activity when added to a Δ*tasA* mutant (Chai et al., 2013), again this was not found to be the case in our hands (Erskine et al., 2018). The disparity between our results and those reported could be due to working with different forms of TasA. This could be impacted by purification conditions (pH, protein concentration and buffer properties) and the storage and handling of the protein used in the assays. It has been reported that TasA adopts multiple oligomeric forms *in vitro*, but TasA undergoes a change to only one of these forms within the biofilm to form fibres (Diehl et al., 2018).

4.1.6. Analysis of the TapA signal sequence

Using the online SignalP 4.1 server to analyse the TapA protein sequence *in silico* it is predicted that TapA has a signal sequence which is cleaved between amino acids 43 and 44 (Petersen et al., 2011). The *B. subtilis* genome encodes

5 type I signal peptidases (Tjalsma et al., 1998). Most signal peptides in *B. subtilis* are cleaved by the action of the serine-lysine type signal peptidases, namely, SipS and SipT (Tjalsma et al., 1999). However, it is SipW that is reported to cleave the signal sequences of both TasA and TapA (Stöver and Driks, 1999a, c). The *tapA*, *sipW* and *tasA* genes are co-located in an operon. The SipW signal peptidase is of the serine-histidine dyad type (Tjalsma et al., 2000b). At the -1 and -3 positions relative to the start site TapA has a threonine and alanine residue respectively which fits with the idea that you need small uncharged amino acids at these positions (Fikes et al., 1990; Shen et al., 1991). It has been reported using *E. coli* as a model system that prolines are not tolerated in the +1 position relative to the cleavage site (the residue that would form the N-terminus of the mature protein) (Barkocy-Gallagher and Bassford, 1992). Furthermore, through global analysis of predicted *B. subtilis* signal peptides it was shown that prolines are never found in the +1 position (Tjalsma et al., 2000a). This work showed that changing alanine at (the predicted) position +1 to proline in the full-length TapA sequence did not alter biofilm morphology and TapA was still processed to lower molecular weight forms (**Figure 32**: B and C). This indicates that TapA still achieves an extracellular location, in the case of $TapA_{A44P}$, as processing is exoprotease-dependent. Signal sequences which are not cleaved can result in a membrane anchored protein (Dalbey and Wickner, 1985; Fikes et al., 1987). It is possible that the predicted TapA signal peptide is actually uncleaved *in vivo* resulting in the anchoring of the protein to the membrane. In this model, the C-terminus is found in the periplasmic space or in association with the cell wall. In support of this hypothesis, fusion of the known TasA signal peptide in place of the first 43 amino acids of the TapA sequence, of the minimal functional unit, to encode

$TasA_{sp}$-$TapA_{44-57}$, could not facilitate wild-type biofilm formation. The discrimination score (*D*) for the TapA signal sequence is 0.527, given by SignalP4.1, it is very close to the cut-off of 0.45 indicating that the probability of it being a signal peptide is relatively low with scores above 0.7 indicating a high confidence of a correct prediction (Petersen et al., 2011). Finally, deletion of *sipW* did not impact the extracellular processing of TapA as judged by an αTapA immunoblot (**Figure 32**C). This is in contradiction to the report that SipW cleaves the signal peptide of TapA (Stöver and Driks, 1999a).

4.1.7. The role of exoproteases in biofilm formation

Exoproteases are known to be produced during biofilm formation (Cairns et al., 2014; Marlow et al., 2014). The findings of the current work, specifically that TapA is processed by secreted proteases, is novel and interesting (**Figure 20**B). It was discovered that even in the absence of secreted proteases *B. subtilis* still formed structurally complex biofilms (**Figure 20**C). In reality, the functional role of proteases in the *B. subtilis* biofilm is an under explored topic. Uncovering the role of exoproteases may require a more sophisticated approach than macroscopic observations of biofilm morphology. One explanation could be that proteolytic cleavage of TapA serves as an additional regulatory role in governing biofilm formation, but that under the artificial conditions tested the impact of perturbing this process (by deleting exoprotease genes) is not readily observable.

4.2. A model for the function of TapA

Given all of the information above and the published information regarding TapA, I propose the following outline for how TapA functions. The TapA protein is transported to the membrane where the signal sequence is cleaved to

generate the mature form of the protein $TapA_{44\text{-}253}$. The work here demonstrates that TapA export does not seem to be SipW-dependent, as TapA is still subject to extracellular degradation when the *sipW* gene is deleted (**Figure 32**C). More importantly, the observed processing of TapA by extracellular proteases (**Figure 20**B) is in support of the hypothesis that TapA is indeed a secreted protein, as predicted *in silico*. TasA fibres then form by the polymerisation of monomeric TasA (mTasA) subunits in a TapA-dependent manner (**Figure 35**A). The mature $TapA_{44\text{-}253}$ protein contains the essential β-$strand_{51\text{-}56}$ which has been the subject of investigation in the current work. It may be that the β-$strand_{51\text{-}56}$ is a nucleation point for the formation of TasA fibres from monomeric subunits. Consistent with this suggestion is the published finding that TapA aides the polymerisation of TasA *in vitro* but that TasA polymerisation is inhibited by $TapA_{\Delta 50\text{-}57}$ in which the genetic region corresponding to the β-strand region has been deleted (Romero et al., 2014).

Our own published findings show that TasA rendered monomeric (mTasA) is subject to exoprotease degradation but that the fibrous form is resistant to proteolysis (Erskine et al., 2018). It may be that in the absence of TapA, TasA remains monomeric and is broken down almost completely, with fibres only forming in the presence of TapA (**Figure 35**B). Therefore, degraded forms of TasA are present in wild-type biofilms because the degradation of TasA is reduced and incomplete. We have found that recombinant TasA will form fibres spontaneously *in vitro* (Erskine et al., 2018). The reason that TasA fibres form *in vivo* may be due to TapA interacting with TasA to keep the localised concentration higher. This is the same reason proposed for TasA formation *in*

vitro since purified recombinant TasA is stored as a high concentration solution in our work.

It was hypothesised, as part of the current work, that TapA function was regulated by proteolysis mediated by exoproteases. Although TapA processing was perturbed in the absence of exoproteases (**Figure 20**B), it was found that unprocessed TapA is still functional (**Figure 20**C). It is not clear whether we are actually detecting the unprocessed form of TapA by αTapA immunoblot or an alternatively processed form of TasA due to the non-specific nature of the antibody. The actual form of the mature TapA protein that is functional is unclear, but to summarise we have shown that unprocessed TapA may resemble $TapA_{44-253}$ (**Figure 35**A). Importantly we also showed that $TapA_{1-57}$ is functional (**Figure 35**A). It could be that the functional form, in WT biofilms, is $TapA_{44-253}$ and once it has carried out its function it is unstable and degraded by exoproteases. Importantly, both $TapA_{44-253}$ and $TapA_{1-57}$ contain the β-$strand_{51-56}$ shown in this work to be key for function (**Figure 28**). At this point, it is not clear whether TapA functions by nucleating/assisting TasA fibre formation or alternatively it could be a chaperone preventing the degradation of mTasA.

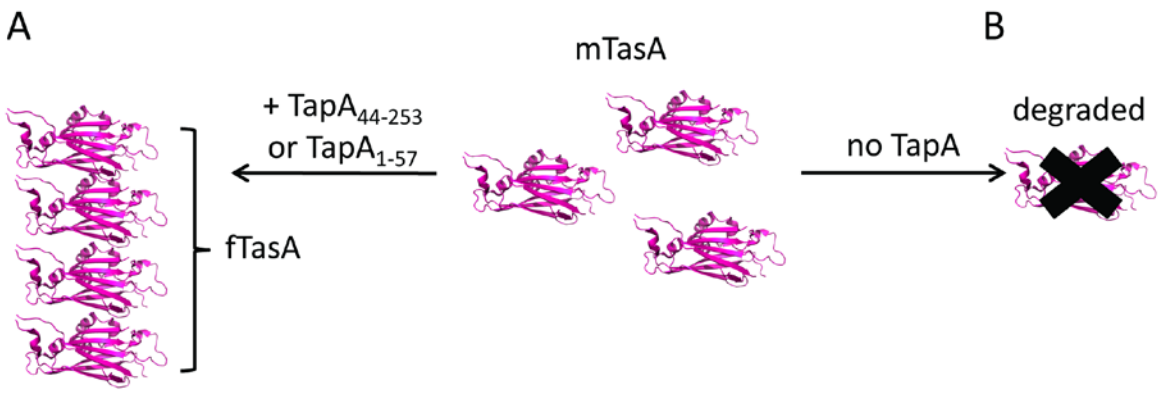


Figure 35: A model for TapA function. (A) In the presence of $TapA_{44-253}$ or $TapA_{1-57}$ the formation of fibrous TasA (fTasA) is supported from monomeric TasA (mTasA) subunits. It is not clear whether TapA functions to nucleate fibre assembly or acts as a chaperone that prevents the degradation of TasA. **(B)** In the absence of TapA, TasA remains monomeric, and is unstable.

Alternative explanations for the function of TapA

Given that there are many unanswered questions as to how TapA functions in the development of complexity in the *B. subtilis* biofilm, I here propose some alternative roles.

The current published model for TapA function describes TapA as anchoring TasA fibres to the cell wall with TasA fibres extending from the surface of the cell into the extracellular environment. There is an additional role proposed for TapA in the assembly of TasA fibres as suggested by a number of different experimental approaches both *in vitro* and *in vivo*. However, we found that the presence of *tapA* was not needed for the biological function of TasA fibres added exogeneously to biofilms, therefore, suggesting that a TapA-dependent attachment point is not critical for the function of TasA (Erskine et al., 2018). Moreover, the TapA and TasA sequences do not include any likely candidates for cell wall anchoring motifs, which normally exist in the sequence context, for Gram-positive bacteria, of LPXTG. As an example, the LPXTG motif is found in

the cell wall anchored (CWA) proteins of *Staphylococcal aureus* which are important for the attachment of cells to a surface in order to initiate biofilm formation (Foster et al., 2013). They are also important for binding of ligands found in the human host (Hobley et al., 2015). Additionally, in *S. aureus* biofilms the surfaces of cells are decorated with fibres formed by monomers called phenol soluble modulins (Periasamy et al., 2012). It is unclear how these fibres are attached to the surface of the cell but this could be by electrostatic association with the cell wall. It could be that an association of TasA and TapA with the cell wall is mediated by electrostatic interactions.

A previous publication (Tjalsma et al., 2000b), describes the growth of *Bacillus subtilis* in TY media and controlled *sipW* expression with a constitutively active promoter. They found that after 4h TasA remained unprocessed even though *sipW* was expressed. The authors suggest that this indicates there is an as yet unidentified factor which is present at the later time points where TasA was observed in a processed form. It is possible that this additional factor is TapA. It should be noted that the authors could not detect TapA in these growth conditions but this could be related to issues with the αTapA antibody, which I have discussed. It is difficult to explain how TapA would become subject to extracellular proteolysis given this model. It could be that TapA is also exported as part of this process or TapA may only reach an extracellular location because of cell lysis. Curli fibres of *E. coli* are the most abundant proteinaceous component in *E. coli* biofilms (and other related species) (Evans and Chapman, 2014). In this case, fibres are formed by the subunit CsgA and the nucleation of fibrils is carried out by CsgB. Nucleation by CsgB is depedent upon CsgF which is associated with the outer membrane from which fibres form (Hobley et al., 2015). CsgF interacts with the Curli specific transport apparatus CsgG (Hobley

et al., 2015). It could be that TapA operates in a dual role by nucleating fibres (like CsgB) but also associating with the outer surface of the membrane (like CsgF) to facilitate TasA export. Alternatively, TapA may act as a chaperone to keep TasA in an unfolded state for Sec-dependent transport. In this case, when the *tapA* gene is absent then TasA is degraded by intracellular proteases.

It may be that the TapA signal peptide is cleaved off by an alternative signal peptidase, other than SipW, with the mature TapA undergoing further processing to generate a secreted peptide consisting of the intervening amino acids from 44-57. It is anticipated that this would appear to be the case for the minimal functional form of $TapA_{1-57}$. This peptide could nucleate the formation of TasA fibres in the extracellular environment. RbmA is an example of a secreted, matrix protein of *Vibrio cholerae* which is needed for the rugosity of biofilms (Fong et al., 2010). It is exported via a Type II secretion system into the extracellular environment where it is thought to function by binding to carbohydrates (Johnson et al., 2014). Alternatively, the TapA peptide may anchor fibres to the cell surface as has been previously postulated (Romero et al., 2011).

Finally, it is possible that the low scoring *in silico* predicted TapA signal peptide is indicative of an uncleaved membrane anchoring domain. In this model TapA is a membrane protein with a single pass N-terminal transmembrane helix and the C-terminus of the protein would be localised externally to the cell exposing it to the protease activity of the extracellular environment. This membrane anchored TapA could function in a similar way to that described previously for TapA by anchoring TasA fibres to the cell surface.

4.3. Future outlook

It remains unclear what the role of the stable core of the TapA protein and the C-terminus is but given that there are many very well conserved features, in the protein core, this indicates it may have a biofilm-independent function. Alternatively, it could be that the protein functions exclusively in biofilm formation and the *tapA* gene is conserved to encode an appropriately processed form of TapA. The alternative explanation may be the most likely given that the minimal functional form of TapA ($TapA_{1-57}$) still conveys rugosity to the biofilm without the presence of the rest of the protein.

The results of this work have shown that exoproteases are not needed for the structural complexity of *B. subtilis* biofilms. The artificial growth conditions implemented in laboratory studies only offer some insight into the real world lifestyle of microorganisms. Therefore, it would be interesting to study the ability of the exoprotease-free strain to form biofilms on plant roots. A number of plant models of *B. subtilis* biofilm formation are available including *Aribidopsis* and Tomato plant root colonization assays (Bais et al., 2004; Chen et al., 2012; Chen et al., 2013). It is possible that exoproteases function in a way to release energy from large biomolecules found in the rhizosphere. This may release amino acids for carbon metabolism and/or as a nitrogen source. The MSgg agar we utilise for biofilm formation assays contains 2 readily available carbon sources in the form of glycerol and glutamic acid, glutamic acid also serves as the sole nitrogen source in MSgg. This may explain the increased expression of exoprotease genes at later time points in biofilm formation experiments (Marlow et al., 2014), as resources are used up. It might be that exoproteases breakdown matrix components to allow for the release of cells when conditions become hostile.

A TapA 3D-structure prediction shows an N-terminal α-helical domain and a β-strand-rich core region (**Figure 9**B). It may be that loop regions connecting these structured domains are vulnerable to proteolysis during the *in vivo* processing of TapA. However, it is notoriously difficult to predict 3D structure from the amino acid sequence of a protein. Experimental investigation of TapA protein structure would be a necessary step in understanding this protein further. Further work may finally confirm the *in vivo* functional form of the protein and this may represent the form which is most suitable for crystallisation studies.

The difficulty in isolating the functional form of TapA arises due to issues with the TapA antibodies used in this study, *i.e.* the novel αTapA that we generated and the TapA anti-serum kindly gifted by Adam Driks at Loyola University. These antibodies do not appear to detect the minimal functional forms of TapA and react non-specifically with the TasA protein. This may be unsurprising given that TasA is the most abundant protein in the biofilm so is present in higher concentration in biofilm lysates. It may be that the rabbit host, in which antibodies were raised, was naturally exposed to *Bacillus subtilis* (also known as the hay *Bacillus* or grass *Bacillus*) and therefore TasA, indirectly, prior to immunisation. To address this, assessment of the pre-immune serum could be tested for specificity against TasA and TapA. A possible solution could be to generate an antibody against the $TapA_{44-57}$ peptide. The key issue that will need to be addressed, is determining the localisation of TapA, and alternative antibodies may help to tackle this. This will hopefully lead to a more informed model as to how TapA functions.

Another method would be to bypass the use of an αTapA antibody. To achieve this goal, tagging the N-terminus of the TapA protein may allow for the isolation

of the functional form of TapA, if it is not cleaved by the action of a signal peptidase. The results presented here cast doubt on the role of SipW as processing the signal peptide of TapA. It should be further investigated whether another signal peptidase performs this role or whether the N-terminus is not cleaved. Finally, it could be determined whether TapA performs a chaperone role and is needed for TasA export by assessing the localisation of TasA (whether it is intracellular or extracellular) in a Δ*tapA* mutant.

Bacillus cereus is a relative of *B. subtilis* and forms biofilms under similar conditions but it has been reported that it does not encode a TapA homologue. Interestingly, the *B. cereus* genome has 2 orthologues of *tasA*, named *tasA* and *calY* (**Figure 36**). In the case of *B. cereus*, TasA was shown to develop more abundant fibres than CalY but neither were essential for biofilm formation in contrast to the critical role of TasA in *B. subtilis* biofilm formation (Caro-Astorga et al., 2014). It may be that *B. cereus* resembles the ancestral system that preceded the *tapA-sipW-tasA* system of *Bacillus subtilis*, although this remains an unexplored hypothesis. In this evolutionary model TapA may have arisen from an extra homolog of TasA and adopted a role as an additional regulatory control in TasA fibre formation *in vivo*. Phylogenetic analysis may be able to infer whether this was the case or not.

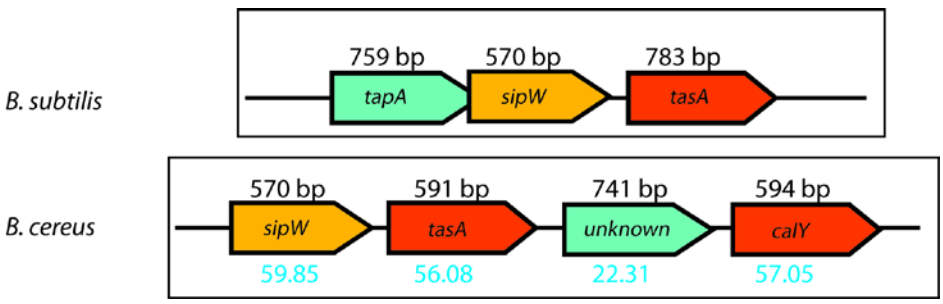


Figure 36: The architecture of the *B. cereus tapA* operon by comparison with the *B. subtilis tapA* operon. The length of the ORFs are shown directly above the gene sequences in black text. *B. cereus* has 2 *tasA* homologs (termed *tasA* and *calY*), as part of this operon.

Between *tasA* and *calY* there is an ORF of unknown function. This ORF has some level of sequence identity with *tapA* (22.31) but was not found to be expressed in the same conditions in which other members of this operon were being expressed, namely when *B. cereus* was grown in LB. The percent sequence identity is shown in blue for *B. cereus* with reference to the *B. subtilis* sequences.

I have discussed in some detail the real world socio-economic importance of biofilms both “good” and “bad”. More recently research groups are now trying to design biofilms by engineering them to have new properties. For example, *E. coli* biofilms have been engineered to create synthetic biological nanomaterials (Nguyen, 2017). It could be that the TasA system of *B. subtilis* offers an alternative model with different properties which may facilitate the creation of novel biological nanomaterials. Furthermore, by understanding biofilm formation in *B. subtilis* it may be possible to engineer improved biocontrol strains with enhanced biofilm forming properties.

5. Bibliography

Aguilar, C., Vlamakis, H., Guzman, A., Losick, R., and Kolter, R. (2010). KinD is a checkpoint protein linking spore formation to extracellular-matrix production in *Bacillus subtilis* biofilms. mBio *1*.
Aguilar, C., Vlamakis, H., Losick, R., and Kolter, R. (2007). Thinking about *Bacillus subtilis* as a multicellular organism. Current opinion in microbiology *10*, 638-643.
Aguilera, A., Souza-Egipsy, V., Gómez, F., and Amils, R. (2007). Development and structure of eukaryotic biofilms in an extreme acidic environment, Río Tinto (SW, Spain). Microbial Ecology *53*, 294-305.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of Molecular Biology *215*, 403-410.
Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research *25*, 3389-3402.
Amati, G., Bisicchia, P., and Galizzi, A. (2004). DegU-P represses expression of the motility fla-che operon in *Bacillus subtilis*. Journal of bacteriology *186*, 6003-6014.
Anwar, A., and Saleemuddin, M. (1998). Alkaline proteases: A review. Bioresource Technology *64*, 175-183.
Arima, K., Kakinuma, A., and Tamura, G. (1968). Surfactin, a crystalline peptidelipid surfactant produced by *Bacillus subtilis*: Isolation, characterization and its inhibition of fibrin clot formation. Biochemical and biophysical research communications *31*, 488-494.
Arnaouteli, S., Ferreira, A.S., Schor, M., Morris, R.J., Bromley, K.M., Jo, J., Cortez, K.L., Sukhodub, T., Prescott, A.R., Dietrich, L.E.P*., et al.* (2017). Bifunctionality of a biofilm matrix protein controlled by redox state. Proceedings of the National Academy of Sciences *114*, E6184-E6191.
Arnaouteli, S., MacPhee, C.E., and Stanley-Wall, N.R. (2016). Just in case it rains: building a hydrophobic biofilm the *Bacillus subtilis* way. Current Opinion in Microbiology *34*, 7-12.
Arnaud, M., Chastanet, A., and Débarbouillé, M. (2004). New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, Gram-positive Bacteria. Applied and Environmental Microbiology *70*, 6887-6891.
Bais, H.P., Fall, R., and Vivanco, J.M. (2004). Biocontrol of *Bacillus subtilis* against infection of Arabidopsis roots by *Pseudomonas syringae* is facilitated by biofilm formation and surfactin production. Plant physiology *134*, 307-319.
Barkocy-Gallagher, G.A., and Bassford, P.J. (1992). Synthesis of precursor maltose-binding protein with proline in the +1 position of the cleavage site interferes with the activity of *Escherichia coli* signal peptidase I *in vivo*. Journal of Biological Chemistry *267*, 1231-1238.
Beauregard, P.B., Chai, Y., Vlamakis, H., Losick, R., and Kolter, R. (2013). *Bacillus subtilis* biofilm induction by plant polysaccharides. Proceedings of the National Academy of Sciences *110*, E1621-E1630.
Bennhold, H. (1922). Eine spezifische Amyloidfarbung wit Kongorot. Meunch Med Wocheschr *69*, 1537-1538.
Blair, J.M.A., Webber, M.A., Baylay, A.J., Ogbolu, D.O., and Piddock, L.J.V. (2014). Molecular mechanisms of antibiotic resistance. Nature Reviews Microbiology *13*, 42.
Bottone, E.J. (2010). *Bacillus cereus*, a Volatile Human Pathogen. Clinical Microbiology Reviews *23*, 382-398.
Branda, S.S., Chu, F., Kearns, D.B., Losick, R., and Kolter, R. (2006). A major protein component of the *Bacillus subtilis* biofilm matrix. Molecular Microbiology *59*, 1229-1238.
Branda, S.S., González-Pastor, J.E., Ben-Yehuda, S., Losick, R., and Kolter, R. (2001). Fruiting body formation by Bacillus subtilis. Proceedings of the National Academy of Sciences *98*, 11621-11626.
Breydo, L., Wu, J.W., and Uversky, V.N. (2012). α-Synuclein misfolding and Parkinson's disease. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease *1822*, 261-285.
Bridier, A., Briandet, R., Thomas, V., and Dubois-Brissonnet, F. (2011). Resistance of bacterial biofilms to disinfectants: a review. Biofouling *27*, 1017-1032.

Britton, R.A., Eichenberger, P., Gonzalez-Pastor, J.E., Fawcett, P., Monson, R., Losick, R., and Grossman, A.D. (2002). Genome-Wide Analysis of the Stationary-Phase Sigma Factor (Sigma-H) Regulon of *Bacillus subtilis*. Journal of Bacteriology *184*, 4881-4890.
Bromley, K.M., Morris, R.J., Hobley, L., Brandani, G., Gillespie, R.M.C., McCluskey, M., Zachariae, U., Marenduzzo, D., Stanley-Wall, N.R., and MacPhee, C.E. (2015). Interfacial self-assembly of a bacterial hydrophobin. Proceedings of the National Academy of Sciences *112*, 5419-5424.
Brown, L., Kessler, A., Cabezas-Sanchez, P., Luque-Garcia, J.L., and Casadevall, A. (2014). Extracellular Vesicles Produced by the Gram-positive Bacterium *Bacillus subtilis* are Disrupted by the Lipopeptide Surfactin. Molecular microbiology *93*, 183-198.
Cairns, L.S., Hobley, L., and Stanley-Wall, N.R. (2014). Biofilm formation by *Bacillus subtilis*: new insights into regulatory strategies and assembly mechanisms. Molecular Microbiology *93*, 587-598.
Cairns, L.S., Martyn, J.E., Bromley, K., and Stanley-Wall, N.R. (2015). An alternate route to phosphorylating DegU of *Bacillus subtilis* using acetyl phosphate. BMC Microbiology *15*, 78.
Caro-Astorga, J., Pérez-García, A., de Vicente, A., and Romero, D. (2014). A genomic region involved in the formation of adhesin fibers in *Bacillus cereus* biofilms. Frontiers in Microbiology *5*, 745.
Chai, L., Romero, D., Kayatekin, C., Akabayov, B., Vlamakis, H., Losick, R., and Kolter, R. (2013). Isolation, Characterization, and Aggregation of a Structured Bacterial Matrix Precursor. Journal of Biological Chemistry *288*, 17559-17568.
Chai, Y., Chu, F., Kolter, R., and Losick, R. (2008). Bistability and Biofilm Formation in *Bacillus subtilis*. Molecular microbiology *67*, 254-263.
Chandki, R., Banthia, P., and Banthia, R. (2011). Biofilms: A microbial home. Journal of Indian Society of Periodontology *15*, 111-114.
Charlton, T.S., De Nys, R., Netting, A., Kumar, N., Hentzer, M., Givskov, M., and Kjelleberg, S. (2000). A novel and sensitive method for the quantification of N-3-oxoacyl homoserine lactones using gas chromatography–mass spectrometry: application to a model bacterial biofilm. Environmental Microbiology *2*, 530-541.
Chen, Y., Cao, S., Chai, Y., Clardy, J., Kolter, R., Guo, J.h., and Losick, R. (2012). A *Bacillus subtilis* sensor kinase involved in triggering biofilm formation on the roots of tomato plants. Molecular Microbiology *85*, 418-430.
Chen, Y., Yan, F., Chai, Y., Liu, H., Kolter, R., Losick, R., and Guo, J.h. (2013). Biocontrol of tomato wilt disease by *Bacillus subtilis* isolates from natural environments depends on conserved genes mediating biofilm formation. Environmental Microbiology *15*, 848-864.
Chu, F., Kearns, D.B., Branda, S.S., Kolter, R., and Losick, R. (2006). Targets of the master regulator of biofilm formation in *Bacillus subtilis*. Molecular Microbiology *59*, 1216-1228.
Chu, F., Kearns, D.B., McLoon, A., Chai, Y., Kolter, R., and Losick, R. (2008). A novel regulatory protein governing biofilm formation in *Bacillus subtilis*. Molecular microbiology *68*, 1117-1127.
Chubukov, V., and Sauer, U. (2014). Environmental Dependence of Stationary-Phase Metabolism in *Bacillus subtilis* and *Escherichia coli*. Applied and Environmental Microbiology *80*, 2901-2909.
Claessen, D., Rozen, D.E., Kuipers, O.P., Søgaard-Andersen, L., and van Wezel, G.P. (2014). Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies. Nature Reviews Microbiology *12*, 115.
Cohn, F. (1877). Untersuchungen uber bacterien. IV. Beitrage zur biologie der Pflanzen *7*, 249-276.
Conn, H.J. (1930). The Identity of *Bacillus Subtilis*. The Journal of Infectious Diseases *46*, 341-350.
Dahl, M., Msadek, T., Kunst, F., and Rapoport, G. (1992). The phosphorylation state of the DegU response regulator acts as a molecular switch allowing either degradative enzyme synthesis or expression of genetic competence in *Bacillus subtilis*. Journal of Biological Chemistry *267*, 14509-14514.

Dalbey, R.E., and Wickner, W. (1985). Leader peptidase catalyzes the release of exported proteins from the outer surface of the *Escherichia coli* plasma membrane. Journal of Biological Chemistry *260*, 15925-15931.
Das, T., Sehar, S., and Manefield, M. (2013). The roles of extracellular DNA in the structural integrity of extracellular polymeric substance and bacterial biofilm development. Environmental Microbiology Reports *5*, 778-786.
Decho, A.W., Visscher, P.T., and Reid, R.P. (2005). Production and cycling of natural microbial exopolymers (EPS) within a marine stromatolite. Palaeogeography, Palaeoclimatology, Palaeoecology *219*, 71-86.
DeLano, W. (2016). The PyMOL Molecular Graphics System. DeLano Scientific; Palo Alto, CA: 2002.
Diehl, A., Roske, Y., Ball, L., Chowdhury, A., Hiller, M., Molière, N., Kramer, R., Stöppler, D., Worth, C.L., Schlegel, B.*, et al.* (2018). Structural changes of TasA in biofilm formation of *Bacillus subtilis*. Proceedings of the National Academy of Sciences *115*, 3237-3242.
Divry, P. (1927). Etude histochimique des plaques seniles. J Belge Neurol Psychiat *27*, 643-657.
Dragoš, A., Kiesewalter, H., Martin, M., Hsu, C.-Y., Hartmann, R., Wechsler, T., Eriksen, C., Brix, S., Drescher, K., Stanley-Wall, N.*, et al.* (2018). Division of Labor during Biofilm Matrix Production. Current Biology, *In press*.
Dragos, A., Kiesewalter, H.T., Martin, M., Hsu, C.-Y., Hartmann, R., Wechsler, T., Drescher, K., Stanley-Wall, N., Kummerli, R., and Kovacs, A.T. (2017). Division of labor during biofilm matrix production. bioRxiv.
Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015). JPred4: a protein secondary structure prediction server. Nucleic Acids Research *43*, W389-W394.
Dubnau, D., and Losick, R. (2006). Bistability in bacteria. Molecular microbiology *61*, 564-572.
Duc, L.H., Hong, H.A., Fairweather, N., Ricca, E., and Cutting, S.M. (2003). Bacterial Spores as Vaccine Vehicles. Infection and Immunity *71*, 2810-2818.
Dufrêne, Y.F. (2015). Sticky microbes: forces in microbial cell adhesion. Trends in microbiology *23*, 376-382.
Duthie, E.S. (1944). The Production of Penicillinase by Organisms of the *Subtilis* Group. British Journal of Experimental Pathology *25*, 96-100.
Earl, A.M., Losick, R., and Kolter, R. (2008). Ecology and genomics of *Bacillus subtilis*. Trends in Microbiology *16*, 269-275.
Elshaghabee, F.M.F., Rokana, N., Gulhane, R.D., Sharma, C., and Panwar, H. (2017). Bacillus As Potential Probiotics: Status, Concerns, and Future Perspectives. Frontiers in Microbiology *8*.
Epstein, A.K., Pokroy, B., Seminara, A., and Aizenberg, J. (2011). Bacterial biofilm shows persistent resistance to liquid wetting and gas penetration. Proceedings of the National Academy of Sciences *108*, 995-1000.
Ereshefsky, M., and Pedroso, M. (2015). Rethinking evolutionary individuality. Proceedings of the National Academy of Sciences of the United States of America *112*, 10126-10132.
Erskine, E., Morris, R.J., Schor, M., Earl, C., Gillespie, R.M.C., Bromley, K., Sukhodub, T., Clark, L., Fyfe, P.K., Serpell, L.C.*, et al.* (2018). Formation of functional, non-amyloidogenic fibres by recombinant *Bacillus subtilis* TasA. Molecular Microbiology, *In press*.
Evans, M.L., and Chapman, M.R. (2014). Curli biogenesis: Order out of disorder. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1843*, 1551-1558.
Fakhry, S., Sorrentini, I., Ricca, E., De Felice, M., and Baccigalupi, L. (2008). Characterization of spore forming Bacilli isolated from the human gastrointestinal tract. Journal of Applied Microbiology *105*, 2178-2186.
Fall, R., Kearns, D.B., and Nguyen, T. (2006). A defined medium to investigate sliding motility in a *Bacillus subtilis* flagella-less mutant. BMC Microbiology *6*, 31.
Feng, L., Wu, Z., and Yu, X. (2013). Quorum sensing in water and wastewater treatment biofilms. Journal of environmental biology *34*, 437.

Fikes, J.D., Bankaitis, V.A., Ryan, J.P., and Bassford, P.J. (1987). Mutational alterations affecting the export competence of a truncated but fully functional maltose-binding protein signal peptide. Journal of Bacteriology *169*, 2345-2351.
Fikes, J.D., Barkocy-Gallagher, G.A., Klapper, D.G., and Bassford, P.J. (1990). Maturation of *Escherichia coli* maltose-binding protein by signal peptidase I in vivo. Sequence requirements for efficient processing and demonstration of an alternate cleavage site. Journal of Biological Chemistry *265*, 3417-3423.
Flemming, H.-C., and Wingender, J. (2010). The biofilm matrix. Nature Reviews Microbiology *8*, 623.
Flemming, H.-C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S.A., and Kjelleberg, S. (2016). Biofilms: an emergent form of bacterial life. Nature Reviews Microbiology *14*, 563.
Fong, J.C.N., Syed, K.A., Klose, K.E., and Yildiz, F.H. (2010). Role of *Vibrio* polysaccharide (*vps*) genes in VPS production, biofilm formation and *Vibrio cholerae* pathogenesis. Microbiology *156*, 2757-2769.
Foster, T.J., Geoghegan, J.A., Ganesh, V.K., and Höök, M. (2013). Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. Nature Reviews Microbiology *12*, 49.
Francolini, I., and Donelli, G. (2010). Prevention and control of biofilm-based medical-device-related infections. FEMS Immunology & Medical Microbiology *59*, 227-238.
Fröls, S. (2013). Archaeal biofilms: widespread and complex (Portland Press Limited).
Fujita, M., González-Pastor, J.E., and Losick, R. (2005). High- and Low-Threshold Genes in the Spo0A Regulon of *Bacillus subtilis*. Journal of Bacteriology *187*, 1357-1368.
Fuqua, W.C., Winans, S.C., and Greenberg, E.P. (1994). Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. Journal of Bacteriology *176*, 269-275.
Gallegos-Monterrosa, R., Kankel, S., Götze, S., Barnett, R., Stallforth, P., and Kovács, Á.T. (2017). *Lysinibacillus fusiformis* M5 Induces Increased Complexity in *Bacillus subtilis* 168 Colony Biofilms via Hypoxanthine. Journal of Bacteriology *199*.
Gallegos-Monterrosa, R., Mhatre, E., and Kovács, Á.T. (2016). Specific *Bacillus subtilis* 168 variants form biofilms on nutrient-rich medium. Microbiology *162*, 1922-1932.
Gautheret, D., and Lambert, A. (2001). Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. Journal of Molecular Biology *313*, 1003-1011.
González-Pastor, J.E. (2011). Cannibalism: a social behavior in sporulating *Bacillus subtilis*. FEMS Microbiology Reviews *35*, 415-424.
González-Pastor, J.E., Hobbs, E.C., and Losick, R. (2003). Cannibalism by sporulating bacteria. Science *301*, 510-513.
Gordon, C., Hodges, N., and Marriott, C. (1988). Antibiotic interaction and diffusion through alginate and exopolysaccharide of cystic fibrosis-derived *Pseudomonas aeruginosa*. Journal of Antimicrobial Chemotherapy *22*, 667-674.
Gordon, C.A., Hodges, N.A., and Marriott, C. (1991). Use of slime dispersants to promote antibiotic penetration through the extracellular polysaccharide of mucoid *Pseudomonas aeruginosa*. Antimicrobial Agents and Chemotherapy *35*, 1258-1260.
Grosberg, R.K., and Strathmann, R.R. (2007). The Evolution of Multicellularity: A Minor Major Transition? Annual Review of Ecology, Evolution, and Systematics *38*, 621-654.
Hall-Stoodley, L., Costerton, J.W., and Stoodley, P. (2004). Bacterial biofilms: from the Natural environment to infectious diseases. Nature Reviews Microbiology *2*, 95.
Hamon, M.A., Stanley, N.R., Britton, R.A., Grossman, A.D., and Lazazzera, B.A. (2004). Identification of AbrB-regulated genes involved in biofilm formation by *Bacillus subtilis*. Molecular Microbiology *52*, 847-860.
Henrichsen, J. (1972). Bacterial surface translocation: a survey and a classification. Bacteriological Reviews *36*, 478-503.

Hense, B.A., Kuttler, C., Müller, J., Rothballer, M., Hartmann, A., and Kreft, J.-U. (2007). Does efficiency sensing unify diffusion and quorum sensing? Nature Reviews Microbiology *5*, 230.
Herron, M.D., Borin, J.M., Boswell, J.C., Walker, J., Knox, C.A., Boyd, M., Rosenzweig, F., and Ratcliff, W.C. (2018). *De novo* origin of multicellularity in response to predation. bioRxiv.
Hillel, D., and Hatfield, J.L. (2005). Encyclopedia of Soils in the Environment, Vol 3 (Elsevier Amsterdam).
Hobley, L., Harkins, C., MacPhee, C.E., and Stanley-Wall, N.R. (2015). Giving structure to the biofilm matrix: an overview of individual strategies and emerging common themes. FEMS Microbiology Reviews *39*, 649-669.
Hobley, L., Ostrowski, A., Rao, F.V., Bromley, K.M., Porter, M., Prescott, A.R., MacPhee, C.E., van Aalten, D.M.F., and Stanley-Wall, N.R. (2013). BslA is a self-assembling bacterial hydrophobin that coats the *Bacillus subtilis* biofilm. Proceedings of the National Academy of Sciences *110*, 13600-13605.
Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M., and Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. Monatshefte für Chemie / Chemical Monthly *125*, 167-188.
Howie, A.J., and Brewer, D.B. (2009). Optical properties of amyloid stained by Congo red: History and mechanisms. Micron *40*, 285-301.
Hughes, A.C., and Kearns, D.B. (2017). Swimming, Swarming and Sliding Motility in *Bacillus subtilis*. Bacillus, 415.
Hughes, V., Jiang, C., and Brun, Y. (2012). *Caulobacter crescentus*. Current biology: CB *22*, R507.
Jang, J., Cho, M., Chun, J.-H., Cho, M.-H., Park, J., Oh, H.-B., Yoo, C.-K., and Rhie, G.-e. (2011). The Poly-γ-d-Glutamic Acid Capsule of *Bacillus anthracis* Enhances Lethal Toxin Activity. Infection and Immunity *79*, 3846-3854.
Jarrell, K.F., and McBride, M.J. (2008). The surprisingly diverse ways that prokaryotes move. Nature Reviews Microbiology *6*, 466.
Johnson, T.L., Fong, J.C., Rule, C., Rogers, A., Yildiz, F.H., and Sandkvist, M. (2014). The Type II Secretion System Delivers Matrix Proteins for Biofilm Formation by *Vibrio cholerae*. Journal of Bacteriology *196*, 4245-4252.
Jones, S.E., Paynich, M.L., Kearns, D.B., and Knight, K.L. (2014). Protection from Intestinal Inflammation by Bacterial Exopolysaccharides. The Journal of Immunology *192*, 4813-4820.
Kaiser, D. (2001). Building a Multicellular Organism. Annual Review of Genetics *35*, 103-123.
Kaiser, D., and Warrick, H. (2011). *Myxococcus xanthus* Swarms Are Driven by Growth and Regulated by a Pacemaker. Journal of Bacteriology *193*, 5898-5904.
Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., and Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. Nature Protocols *7*, 1511.
Kearns, D.B. (2010). A field guide to bacterial swarming motility. Nature reviews Microbiology *8*, 634-644.
Kearns, D.B., Chu, F., Branda, S.S., Kolter, R., and Losick, R. (2005). A master regulator for biofilm formation by Bacillus subtilis. Molecular Microbiology *55*, 739-749.
Kearns, D.B., and Losick, R. (2003). Swarming motility in undomesticated *Bacillus subtilis*. Molecular microbiology *49*, 581-590.
Keim, C.N., Abreu, F., Lins, U., Lins de Barros, H., and Farina, M. (2004a). Cell organization and ultrastructure of a magnetotactic multicellular organism. Journal of Structural Biology *145*, 254-262.
Keim, C.N., Martins, J.L., Abreu, F., Rosado, A.S., de Barros, H.L., Borojevic, R., Lins, U., and Farina, M. (2004b). Multicellular life cycle of magnetotactic prokaryotes. FEMS Microbiology Letters *240*, 203-208.
Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols *10*, 845.
Kesel, S., Grumbein, S., Gümperlein, I., Tallawi, M., Marel, A.-K., Lieleg, O., and Opitz, M. (2016). Direct Comparison of Physical Properties of *Bacillus subtilis* NCIB 3610 and B-1 Biofilms. Applied and Environmental Microbiology *82*, 2424-2432.

Kloepper, J.W., Ryu, C.-M., and Zhang, S. (2004). Induced Systemic Resistance and Promotion of Plant Growth by *Bacillus spp*. Phytopathology *94*, 1259-1266.
Kobayashi, K., and Iwano, M. (2012). BslA(YuaB) forms a hydrophobic layer on the surface of *Bacillus subtilis* biofilms. Molecular Microbiology *85*, 51-66.
Kolodkin-Gal, I., Romero, D., Cao, S., Clardy, J., Kolter, R., and Losick, R. (2010). D-amino acids trigger biofilm disassembly. Science *328*, 627-629.
Konkol, M.A., Blair, K.M., and Kearns, D.B. (2013). Plasmid-Encoded ComI Inhibits Competence in the Ancestral 3610 Strain of *Bacillus subtilis*. Journal of Bacteriology *195*, 4085-4093.
Koo, B.-M., Kritikos, G., Farelli, J.D., Todor, H., Tong, K., Kimsey, H., Wapinski, I., Galardini, M., Cabal, A., Peters, J.M.*, et al.* (2017a). Construction and Analysis of Two Genome-Scale Deletion Libraries for *Bacillus subtilis*. Cell Systems *4*, 291-305.e297.
Koo, H., Allan, R.N., Howlin, R.P., Stoodley, P., and Hall-Stoodley, L. (2017b). Targeting microbial biofilms: current and prospective therapeutic strategies. Nature Reviews Microbiology *15*, 740.
Körstgens, V., Flemming, H.-C., Wingender, J., and Borchard, W. (2001). Influence of calcium ions on the mechanical properties of a model biofilm of mucoid *Pseudomonas aeruginosa*. Water Science and Technology *43*, 49-57.
Kumar, K., Mella-Herrera, R.A., and Golden, J.W. (2010). Cyanobacterial Heterocysts. Cold Spring Harbor Perspectives in Biology *2*, a000315.
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S.*, et al.* (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. Nature *390*, 249.
Laemmli, U.K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature *227*, 680.
Lam, J., Chan, R., Lam, K., and Costerton, J.W. (1980). Production of mucoid microcolonies by *Pseudomonas aeruginosa* within infected lungs in cystic fibrosis. Infection and Immunity *28*, 546-556.
Lamour, G., Nassar, R., Chan, P.H., Bozkurt, G., Li, J., Bui, J.M., Yip, C.K., Mayor, T., Li, H., and Wu, H. (2017). Mapping the broad structural and mechanical properties of amyloid fibrils. Biophysical journal *112*, 584-594.
Lane, N. (2015). The unseen world: reflections on Leeuwenhoek (1677) 'Concerning little animals'. Philosophical Transactions of the Royal Society B: Biological Sciences *370*, 20140344.
Lane, N., and Martin, W. (2010). The energetics of genome complexity. Nature *467*, 929.
LeChevallier, M.W., Babcock, T.M., and Lee, R.G. (1987). Examination and characterization of distribution system biofilms. Applied and Environmental Microbiology *53*, 2714-2724.
Leiman, S.A., May, J.M., Lebar, M.D., Kahne, D., Kolter, R., and Losick, R. (2013). d-Amino Acids Indirectly Inhibit Biofilm Formation in *Bacillus subtilis* by Interfering with Protein Synthesis. Journal of Bacteriology *195*, 5391-5395.
Leiman, S.A., Richardson, C., Foulston, L., Elsholz, A.K.W., First, E.A., and Losick, R. (2015). Identification and Characterization of Mutations Conferring Resistance to d-Amino Acids in *Bacillus subtilis*. Journal of Bacteriology *197*, 1632-1639.
Lesnik, E.A., Sampath, R., Levene, H.B., Henderson, T.J., McNeil, J.A., and Ecker, D.J. (2001). Prediction of rho-independent transcriptional terminators in *Escherichia coli*. Nucleic Acids Research *29*, 3583-3594.
Liu, S.-B., Chen, X.-L., He, H.-L., Zhang, X.-Y., Xie, B.-B., Yu, Y., Chen, B., Zhou, B.-C., and Zhang, Y.-Z. (2013). Structure and Ecological Roles of a Novel Exopolysaccharide from the Arctic Sea Ice Bacterium *Pseudoalteromonas sp.* Strain SM20310. Applied and Environmental Microbiology *79*, 224-230.
Liu, W.-T., Yang, Y.-L., Xu, Y., Lamsa, A., Haste, N.M., Yang, J.Y., Ng, J., Gonzalez, D., Ellermeier, C.D., Straight, P.D.*, et al.* (2010). Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of *Bacillus subtilis*. Proceedings of the National Academy of Sciences *107*, 16286-16290.

Lopez, D., Vlamakis, H., and Kolter, R. (2009). Generation of multiple cell types in *Bacillus subtilis*. FEMS Microbiology Reviews *33*, 152-163.
López, D., Vlamakis, H., Losick, R., and Kolter, R. (2009). Cannibalism enhances biofilm development in *Bacillus subtilis*. Molecular Microbiology *74*, 609-618.
Lu, J., Tappel, R.C., and Nomura, C.T. (2009). Mini-review: biosynthesis of poly (hydroxyalkanoates). Journal of Macromolecular Science, Part C: Polymer Reviews *49*, 226-248.
Ma, J., Peng, J., Wang, S., and Xu, J. (2012). A conditional neural fields model for protein threading. Bioinformatics *28*, i59-i66.
Ma, J., Wang, S., Zhao, F., and Xu, J. (2013). Protein threading using context-specific alignment potential. Bioinformatics *29*, i257-i265.
Macke, T.J., Ecker, D.J., Gutell, R.R., Gautheret, D., Case, D.A., and Sampath, R. (2001). RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Research *29*, 4724-4735.
MacPhee, C.E., and Dobson, C.M. (2000). Formation of Mixed Fibrils Demonstrates the Generic Nature and Potential Utility of Amyloid Nanostructures. Journal of the American Chemical Society *122*, 12707-12713.
Mäder, U., Schmeisky, A.G., Flórez, L.A., and Stülke, J. (2011). Subti Wiki—a comprehensive community resource for the model organism *Bacillus subtilis*. Nucleic acids research *40*, D1278-D1287.
Madsen, J.S., Burmølle, M., Hansen, L.H., and Sørensen, S.J. (2012). The interconnection between biofilm formation and horizontal gene transfer. FEMS Immunology & Medical Microbiology *65*, 183-195.
Marlow, V.L., Cianfanelli, F.R., Porter, M., Cairns, L.S., Dale, J.K., and Stanley-Wall, N.R. (2014). The prevalence and origin of exoprotease-producing cells in the *Bacillus subtilis* biofilm. Microbiology *160*, 56-66.
Mattick, J.S. (2002). Type IV Pili and Twitching Motility. Annual Review of Microbiology *56*, 289-314.
Mayer, C., Moritz, R., Kirschner, C., Borchard, W., Maibaum, R., Wingender, J., and Flemming, H.-C. (1999). The role of intermolecular interactions: studies on model systems for bacterial biofilms. International Journal of Biological Macromolecules *26*, 3-16.
McKenney, P.T., Driks, A., and Eichenberger, P. (2012). The *Bacillus subtilis* endospore: assembly and functions of the multilayered coat. Nature Reviews Microbiology *11*, 33.
Métivier, R., Bourven, I., Labanowski, J., and Guibaud, G. (2013). Interaction of erythromycin ethylsuccinate and acetaminophen with protein fraction of extracellular polymeric substances (EPS) from various bacterial aggregates. Environmental Science and Pollution Research *20*, 7275-7285.
Michna, R.H., Commichau, F.M., Tödter, D., Zschiedrich, C.P., and Stülke, J. (2013). Subti Wiki–a database for the model organism *Bacillus subtilis* that links pathway, interaction and expression information. Nucleic acids research *42*, D692-D698.
Mielich-Süss, B., and Lopez, D. (2015). Molecular mechanisms involved in *Bacillus subtilis* biofilm formation. Environmental microbiology *17*, 555-565.
Moayeri, M., Leppla, S.H., Vrentas, C., Pomerantsev, A.P., and Liu, S. (2015). Anthrax Pathogenesis. Annual Review of Microbiology *69*, 185-208.
Monds, R.D., and O'Toole, G.A. (2009). The developmental model of microbial biofilms: ten years of a paradigm up for review. Trends in Microbiology *17*, 73-87.
Morikawa, M., Kagihiro, S., Haruki, M., Takano, K., Branda, S., Kolter, R., and Kanaya, S. (2006). Biofilm formation by a *Bacillus subtilis* strain that produces γ-polyglutamate. Microbiology *152*, 2801-2807.
Msadek, T. (1999). When the going gets tough: survival strategies and environmental signaling networks in *Bacillus subtilis*. Trends in Microbiology *7*, 201-207.
Nagórska, K., Bikowski, M., and Obuchowski, M. (2007). Multicellular behaviour and production of a wide variety of toxic substances support usage of *Bacillus subtilis* as a powerful biocontrol agent. Acta Biochimica Polonica-english edition- *54*, 495.

Nealson, K., and Hastings, J.W. (1979). Bacterial bioluminescence: its control and ecological significance. Microbiological reviews *43*, 496.
Nguyen, P.Q. (2017). Synthetic biology engineering of biofilms as nanomaterials factories. Biochemical Society Transactions *45*, 585-597.
Nichols, W.W., Evans, M.J., Slack, M.P., and Walmsley, H.L. (1989). The penetration of antibiotics into aggregates of mucoid and non-mucoid *Pseudomonas aeruginosa*. Microbiology *135*, 1291-1303.
Nicolas, P., Mäder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S*., et al.* (2012). Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in *Bacillus subtilis*. Science *335*, 1103-1106.
Nijland, R., Burgess, J.G., Errington, J., and Veening, J.-W. (2010). Transformation of Environmental *Bacillus subtilis* Isolates by Transiently Inducing Genetic Competence. PLOS ONE *5*, e9724.
Nutman, A.P., Bennett, V.C., Friend, C.R.L., Van Kranendonk, M.J., and Chivas, A.R. (2016). Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. Nature *537*, 535.
O'Donnell, A.G., Norris, J.R., Berkeley, R.C.W., Claus, D., Kaneko, T., Logan, N.A., and Nozaki, R. (1980). Characterization of *Bacillus subtilis*, *Bacillus pumilus*, *Bacillus licheniformis*, and *Bacillus amyloliquefaciens* by Pyrolysis Gas-Liquid Chromatography, Deoxyribonucleic Acid-Deoxyribonucleic Acid Hybridization, Biochemical Tests, and API Systems. International Journal of Systematic and Evolutionary Microbiology *30*, 448-459.
Ogura, M., Yamaguchi, H., Yoshida, K.-i., Fujita, Y., and Tanaka, T. (2001). DNA microarray analysis of *Bacillus subtilis* DegU, ComA and PhoP regulons: an approach to comprehensive analysis of *B. subtilis* two-component regulatory systems. Nucleic Acids Research *29*, 3804-3813.
Okshevsky, M., Regina, V.R., and Meyer, R.L. (2015). Extracellular DNA as a target for biofilm control. Current Opinion in Biotechnology *33*, 73-80.
Ordax, M., Marco-Noales, E., López, M.M., and Biosca, E.G. (2010). Exopolysaccharides favor the survival of *Erwinia amylovora* under copper stress through different strategies. Research in Microbiology *161*, 549-555.
Ostrowski, A., Mehert, A., Prescott, A., Kiley, T.B., and Stanley-Wall, N.R. (2011). YuaB Functions Synergistically with the Exopolysaccharide and TasA Amyloid Fibers To Allow Biofilm Formation by *Bacillus subtilis*. Journal of Bacteriology *193*, 4821-4831.
Patrick, J.E., and Kearns, D.B. (2008). MinJ (YvjD) is a topological determinant of cell division in *Bacillus subtilis*. Molecular Microbiology *70*, 1166-1179.
Peng, J., and Xu, J. (2010). Low-homology protein threading. Bioinformatics *26*, i294-i300.
Peng, J., and Xu, J. (2011a). A multiple-template approach to protein threading. Proteins *79*, 1930-1939.
Peng, J., and Xu, J. (2011b). RaptorX: exploiting structure information for protein alignment by statistical inference. Proteins *79*, 161-171.
Periasamy, S., Joo, H.-S., Duong, A.C., Bach, T.-H.L., Tan, V.Y., Chatterjee, S.S., Cheung, G.Y.C., and Otto, M. (2012). How *Staphylococcus aureus* biofilms develop their characteristic structure. Proceedings of the National Academy of Sciences *109*, 1281-1286.
Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth *8*, 785-786.
Pohl, S., Bhavsar, G., Hulme, J., Bloor, A.E., Misirli, G., Leckenby, M.W., Radford, D.S., Smith, W., Wipat, A., and Williamson, E.D. (2013). Proteomic analysis of *Bacillus subtilis* strains engineered for improved production of heterologous proteins. Proteomics *13*, 3298-3308.
Priest, F.G., Goodfellow, M., and Todd, C. (1988). A numerical classification of the genus *Bacillus*. Microbiology *134*, 1847-1882.
Qin, Z., Ou, Y., Yang, L., Zhu, Y., Tolker-Nielsen, T., Molin, S., and Qu, D. (2007). Role of autolysin-mediated DNA release in biofilm formation of *Staphylococcus epidermidis*. Microbiology *153*, 2083-2092.

Randrianjatovo-Gbalou, I., Rouquette, P., Lefebvre, D., Girbal-Neuhauser, E., and Marcato-Romain, C.E. (2017). *In situ* analysis of *Bacillus licheniformis* biofilms: amyloid-like polymers and eDNA are involved in the adherence and aggregation of the extracellular matrix. Journal of Applied Microbiology *122*, 1262-1274.
Roberts, M.S., and Cohan, F.M. (1995). Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. Evolution, 1081-1094.
Romero, D., Aguilar, C., Losick, R., and Kolter, R. (2010). Amyloid fibers provide structural integrity to *Bacillus subtilis* biofilms. Proceedings of the National Academy of Sciences *107*, 2230-2234.
Romero, D., Vlamakis, H., Losick, R., and Kolter, R. (2011). An accessory protein required for anchoring and assembly of amyloid fibres in B. subtilis biofilms. Molecular Microbiology *80*, 1155-1168.
Romero, D., Vlamakis, H., Losick, R., and Kolter, R. (2014). Functional Analysis of the Accessory Protein TapA in *Bacillus subtilis* Amyloid Fiber Assembly. Journal of Bacteriology *196*, 1505-1513.
Roux, D., Cywes-Bentley, C., Zhang, Y.-F., Pons, S., Konkol, M., Kearns, D.B., Little, D.J., Howell, P.L., Skurnik, D., and Pier, G.B. (2015). Identification of Poly-N-acetylglucosamine as a Major Polysaccharide Component of the *Bacillus subtilis* Biofilm Matrix. The Journal of Biological Chemistry *290*, 19261-19272.
Saeed, S.M., and Fine, G. (1967). Thioflavin-T for Amyloid Detection. American Journal of Clinical Pathology *47*, 588-593.
Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H., and Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science *230*, 1350-1354.
Sarkar, S., and Pires, M.M. (2015). d-Amino Acids Do Not Inhibit Biofilm Formation in *Staphylococcus aureus*. PLOS ONE *10*, e0117613.
Saville, R.M., Rakshe, S., Haagensen, J.A.J., Shukla, S., and Spormann, A.M. (2011). Energy-Dependent Stability of *Shewanella oneidensis* MR-1 Biofilms. Journal of Bacteriology *193*, 3257-3264.
Schor, M., Reid, J.L., MacPhee, C.E., and Stanley-Wall, N.R. (2016). The Diverse Structures and Functions of Surfactant Proteins. Trends in Biochemical Sciences *41*, 610-620.
Segev, E., Tellez, A., Vlamakis, H., and Kolter, R. (2015). Morphological Heterogeneity and Attachment of *Phaeobacter inhibens*. PLOS ONE *10*, e0141300.
Sehar, S., and Naz, I. (2016). Role of the Biofilms in Wastewater Treatment. In Microbial Biofilms - Importance and Applications, D. Dhanasekaran, and N. Thajuddin, eds. (Rijeka: InTech), p. Ch. 07.
Selkoe, D.J., and Hardy, J. (2016). The amyloid hypothesis of Alzheimer's disease at 25 years. EMBO Molecular Medicine *8*, 595-608.
Seper, A., Fengler, V.H., Roier, S., Wolinski, H., Kohlwein, S.D., Bishop, A.L., Camilli, A., Reidl, J., and Schild, S. (2011). Extracellular nucleases and extracellular DNA play important roles in *Vibrio cholerae* biofilm formation. Molecular microbiology *82*, 1015-1037.
Seviour, T., Hansen, S.H., Yang, L., Yau, Y.H., Wang, V.B., Stenvang, M.R., Christiansen, G., Marsili, E., Givskov, M., Chen, Y.*, et al.* (2015). Functional Amyloids Keep Quorum-sensing Molecules in Check. Journal of Biological Chemistry *290*, 6457-6469.
Shapiro, J.A. (1988). Bacteria as multicellular organisms. Scientific American *258*, 82-89.
Shemesh, M., and Chai, Y. (2013). A Combination of Glycerol and Manganese Promotes Biofilm Formation in *Bacillus subtilis* via Histidine Kinase KinD Signaling. Journal of Bacteriology *195*, 2747-2754.
Shen, L.M., Lee, J.I., Cheng, S., Jutte, H., Kuhn, A., and Dalbey, R.E. (1991). Use of site-directed mutagenesis to define the limits of sequence variation tolerated for processing of the M13 procoat protein by the *Escherichia coli* leader peptidase. Biochemistry *30*, 11775-11781.
Shih, I.-L., and Van, Y.-T. (2001). The production of poly-(γ-glutamic acid) from microorganisms and its various applications. Bioresource Technology *79*, 207-225.

Shrout, J.D. (2015). A fantastic voyage for sliding bacteria. Trends in microbiology *23*, 244-246.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J.*, et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol *7*.
Skerker, J.M., and Laub, M.T. (2004). Cell-cycle progression and the generation of asymmetry in *Caulobacter crescentus*. Nature Reviews Microbiology *2*, 325.
Sliusarenko, O., Zusman, D.R., and Oster, G. (2007). The Motors Powering A-Motility in *Myxococcus xanthus* Are Distributed along the Cell Body. Journal of Bacteriology *189*, 7920-7921.
Srinivasan, S., Vladescu, I.D., Koehler, S.A., Wang, X., Mani, M., and Rubinstein, S.M. (2018). Matrix Production and Sporulation in *Bacillus subtilis* Biofilms Localize to Propagating Wave Fronts. Biophysical Journal *114*, 1490-1498.
Stanley-Wall, N.R., and MacPhee, C.E. (2015). Connecting the dots between bacterial biofilms and ice cream. Physical Biology *12*, 063001.
Stanley, N.R., and Lazazzera, B.A. (2005). Defining the genetic differences between wild and domestic strains of Bacillus subtilis that affect poly-γ-dl-glutamic acid production and biofilm formation. Molecular Microbiology *57*, 1143-1158.
Stein, T. (2005). *Bacillus subtilis* antibiotics: structures, syntheses and specific functions. Molecular Microbiology *56*, 845-857.
Stewart, P.S., and William Costerton, J. (2001). Antibiotic resistance of bacteria in biofilms. The Lancet *358*, 135-138.
Stöver, A.G., and Driks, A. (1999a). Control of Synthesis and Secretion of the *Bacillus subtilis* Protein YqxM. Journal of Bacteriology *181*, 7065-7069.
Stöver, A.G., and Driks, A. (1999b). Regulation of Synthesis of the *Bacillus subtilis* Transition-Phase, Spore-Associated Antibacterial Protein TasA. Journal of Bacteriology *181*, 5476-5481.
Stöver, A.G., and Driks, A. (1999c). Secretion, Localization, and Antibacterial Activity of TasA, a *Bacillus subtilis* Spore-Associated Protein. Journal of Bacteriology *181*, 1664-1672.
Studier, F.W. (2005). Protein production by auto-induction in high-density shaking cultures. Protein expression and purification *41*, 207-234.
Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of Molecular Biology *189*, 113-130.
Szklarczyk, R., and Heringa, J. (2004). Tracking repeats using significance and transitivity. Bioinformatics *20*, i311-i317.
Tanaka, T., Kawata, M., and Mukai, K. (1991). Altered phosphorylation of *Bacillus subtilis* DegU caused by single amino acid changes in DegS. Journal of bacteriology *173*, 5507-5515.
Terra, R., Stanley-Wall, N.R., Cao, G., and Lazazzera, B.A. (2012). Identification of *Bacillus subtilis* SipW as a Bifunctional Signal Peptidase That Controls Surface-Adhered Biofilm Formation. Journal of Bacteriology *194*, 2781-2790.
Tjalsma, H., Bolhuis, A., Jongbloed, J.D.H., Bron, S., and van Dijl, J.M. (2000a). Signal Peptide-Dependent Protein Transport in *Bacillus subtilis:* a Genome-Based Survey of the Secretome. Microbiology and Molecular Biology Reviews *64*, 515-547.
Tjalsma, H., Bolhuis, A., van Roosmalen, M.L., Wiegert, T., Schumann, W., Broekhuizen, C.P., Quax, W.J., Venema, G., Bron, S., and van Dijl, J.M. (1998). Functional analysis of the secretory precursor processing machinery of *Bacillus subtilis*: identification of a eubacterial homolog of archaeal and eukaryotic signal peptidases. Genes & Development *12*, 2318-2331.
Tjalsma, H., Stöver, A.G., Driks, A., Venema, G., Bron, S., and van Dijl, J.M. (2000b). Conserved Serine and Histidine Residues Are Critical for Activity of the ER-type Signal Peptidase SipW of *Bacillus subtilis*. Journal of Biological Chemistry *275*, 25102-25108.
Tjalsma, H., van den Dolder, J., Meijer, W.J.J., Venema, G., Bron, S., and van Dijl, J.M. (1999). The Plasmid-Encoded Signal Peptidase SipP Can Functionally Replace the Major Signal Peptidases SipS and SipT of *Bacillus subtilis*. Journal of Bacteriology *181*, 2448-2454.
Tomasz, A. (1965). Control of the competent state in *Pneumococcus* by a hormone-like cell product: an example for a new type of regulatory mechanism in bacteria. Nature *208*, 155-159.

Tuttle, M.D., Comellas, G., Nieuwkoop, A.J., Covell, D.J., Berthold, D.A., Kloepper, K.D., Courtney, J.M., Kim, J.K., Barclay, A.M., Kendall, A.*, et al.* (2016). Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. Nature Structural &Amp; Molecular Biology *23*, 409.
Van Acker, H., Van Dijck, P., and Coenye, T. (2014). Molecular mechanisms of antimicrobial tolerance and resistance in bacterial and fungal biofilms. Trends in Microbiology *22*, 326-333.
van Dijl, J., and Hecker, M. (2013). *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. Microbial Cell Factories *12*, 3.
van Gestel, J., Vlamakis, H., and Kolter, R. (2015). From Cell Differentiation to Cell Collectives: *Bacillus subtilis* Uses Division of Labor to Migrate. PLOS Biology *13*, e1002141.
Verhamme, D.T., Kiley, T.B., and Stanley-Wall, N.R. (2007). DegU co-ordinates multicellular behaviour exhibited by *Bacillus subtilis*. Molecular microbiology *65*, 554-568.
Verhamme, D.T., Murray, E.J., and Stanley-Wall, N.R. (2009). DegU and Spo0A Jointly Control Transcription of Two Loci Required for Complex Colony Development by *Bacillus subtilis*. Journal of Bacteriology *191*, 100-108.
Vilain, S., Pretorius, J.M., Theron, J., and Brözel, V.S. (2009). DNA as an Adhesin: *Bacillus cereus* Requires Extracellular DNA To Form Biofilms. Applied and Environmental Microbiology *75*, 2861-2868.
Vlamakis, H., Aguilar, C., Losick, R., and Kolter, R. (2008). Control of cell fate by the formation of an architecturally complex bacterial community. Genes & development *22*, 945-953.
Vlamakis, H., Chai, Y., Beauregard, P., Losick, R., and Kolter, R. (2013). Sticking together: building a biofilm the *Bacillus subtilis* way. Nature reviews Microbiology *11*, 157-168.
Wang, Z.-W., and Chen, S. (2009). Potential of biofilm-based biofuel production. Applied Microbiology and Biotechnology *83*, 1-18.
Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.
Whitchurch, C.B., Tolker-Nielsen, T., Ragas, P.C., and Mattick, J.S. (2002). Extracellular DNA Required for Bacterial Biofilm Formation. Science *295*, 1487-1487.
Whiteley, M., Diggle, S.P., and Greenberg, E.P. (2017). Progress in and promise of bacterial quorum sensing research. Nature *551*, 313.
Wilking, J.N., Zaburdaev, V., De Volder, M., Losick, R., Brenner, M.P., and Weitz, D.A. (2013). Liquid transport facilitated by channels in *Bacillus subtilis* biofilms. Proceedings of the National Academy of Sciences of the United States of America *110*, 848-852.
Wolcott, R., Costerton, J.W., Raoult, D., and Cutler, S.J. (2013). The polymicrobial nature of biofilm infection. Clinical Microbiology and Infection *19*, 107-112.
Yamashita, A., Singh, S.K., Kawate, T., Jin, Y., and Gouaux, E. (2005). Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. Nature *437*, 215.
Yang, M., Dutta, C., and Tiwari, A. (2015). Disulfide-Bond Scrambling Promotes Amorphous Aggregates in Lysozyme and Bovine Serum Albumin. The Journal of Physical Chemistry B *119*, 3969-3981.
Yilmaz, M., Soran, H., and Beyatli, Y. (2006). Antimicrobial activities of some *Bacillus spp.* strains isolated from the soil. Microbiological Research *161*, 127-131.
Zafra, O., Lamprecht-Grandío, M., de Figueras, C.G., and González-Pastor, J.E. (2012). Extracellular DNA Release by Undomesticated *Bacillus subtilis* Is Regulated by Early Competence. PLOS ONE *7*, e48716.
Zeigler, D.R., Prágai, Z., Rodriguez, S., Chevreux, B., Muffler, A., Albert, T., Bai, R., Wyss, M., and Perkins, J.B. (2008). The Origins of 168, W23, and Other *Bacillus subtilis* Legacy Strains. Journal of Bacteriology *190*, 6983-6995.
Zobell, C.E., and Allen, E.C. (1935). The significance of marine bacteria in the fouling of submerged surfaces. Journal of bacteriology *29*, 239.

6. Appendix:

Supplemetary information and publications.

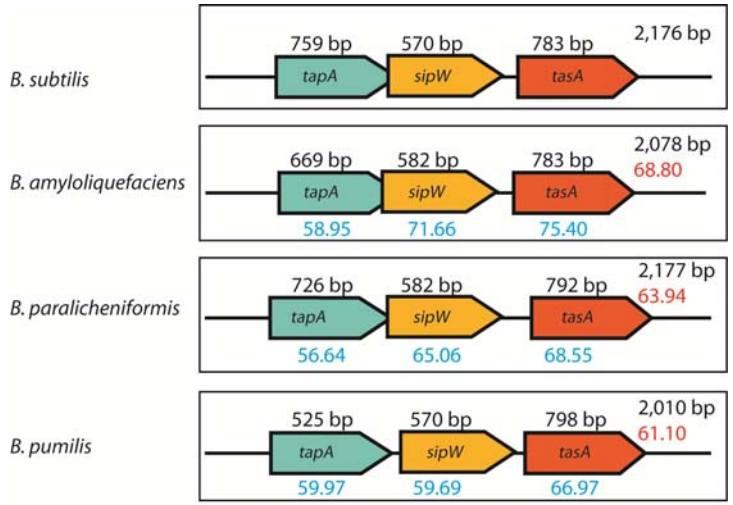


Figure S1: The *tapA* operon is conserved between selected *Bacillus* species. The percentage sequence identity is shown for the *tapA* operon (**red**) for each of the *Bacillus* species analysed in relation to the *Bacillus subtilis* DNA sequence. The percentage identity for each of the genes within the operon, again by comparison with *B. subtilis*, is shown in **blue**. The end of the *tapA* ORF and the start of the *sipW* ORF overlaps by 17 bp, 29 bp and 4 bp in *B. subtilis*, *B. amyloliquefaciens* and *B. paralicheniformis*, respectively. There is a gap of 51 bp between *tapA* and *sipW* in *B. pumilis*. The gap between *tapA* and *tasA* is 63 bp, 64 bp, 72 bp and 57 bp in *B. subtilis*, *B. amyloliquefaciens*, *B. paralicheniformis* and *B. pumilis*, respectively. Accession numbers: *Bacillus subtilis* CP020102.1, *Bacillus amyloliquefaciens* FZB42: CP000560, *Bacillus paralicheniformis*: MIZE01000002.1 and *Bacillus pumilis* SAFR2: CP000813.4.

	TapA				SipW			TasA			
Origin	**Full-length**	**Mature form length**	**Molecular weight (kDa)**	**Percentage identity**	**Full-length**	**Molecular weight (kDa)**	**Percentage identity**	**Full-length**	**Mature form length**	**Molecular weight (kDa)**	**Percentage identity**
B. subtilis	253	210	24.18	-	190	20.68	-	261	234	25.74	-
B. amyloliquefaciens	223	188	21.66	48.89	194	20.8	73.16	261	234	25.56	83.52
B. paralicheniformis	242	199	22.51	38.31	194	21.02	64.74	264	237	26.25	71.21
B. pumilis	175	146	16.68	42	190	20.73	48.66	266	239	26.56	62.17

Table S1: The details of the proteins encoded by the *tapA* operon from the *Bacillus* species selected. Percentage identity is calculated with respect to the amino acid sequences of the *B. subtilis* proteins.

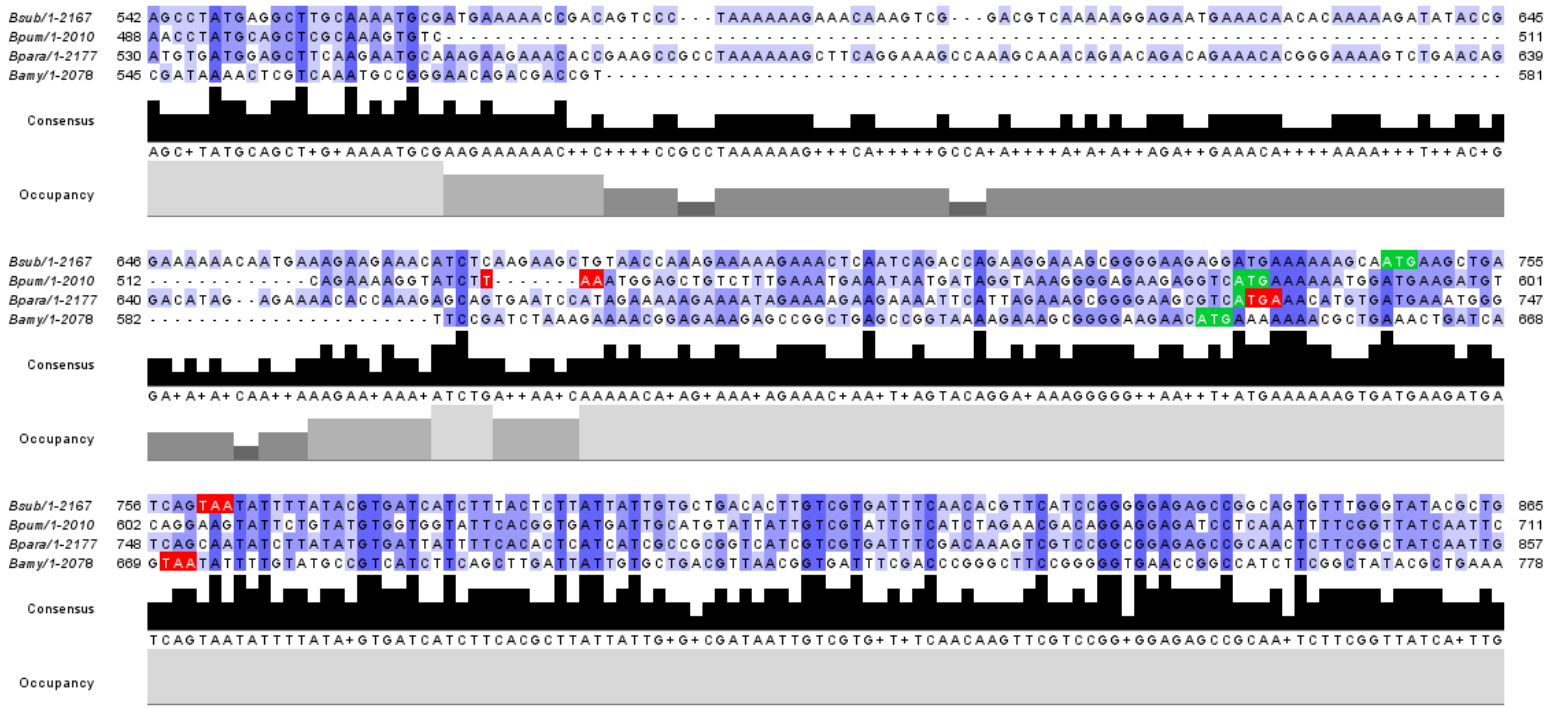


Figure S2: A multiple sequence alignment of *tapA* operons. The multiple sequence alignment was carried out using Clustal Omega (with default settings) and visualised using the Jalview 2 workbench. Colouring in purple is based on percent identity. Highlighted in green is the start codon for *sipW* and in red is the stop codon for the *tapA* ORF.

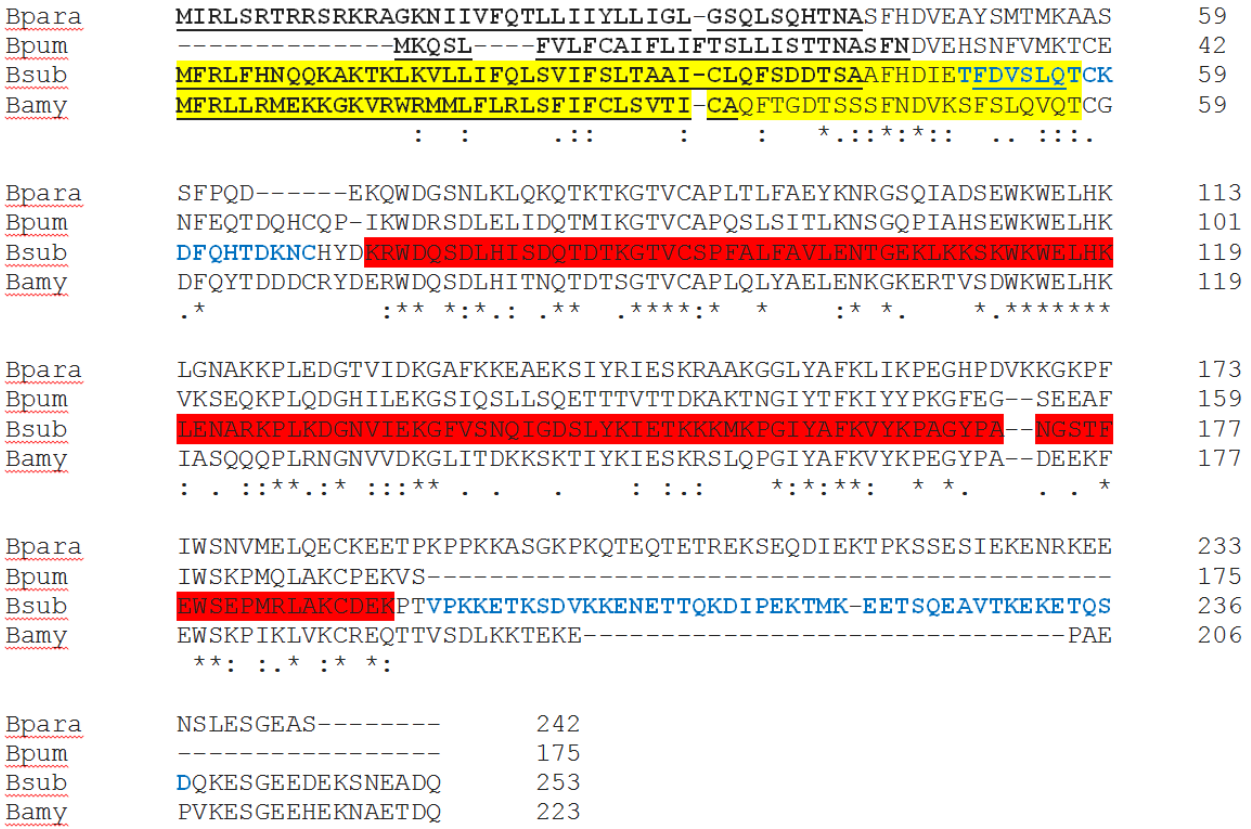


Figure S3: Multiple sequence alignment of TapA. The predicted signal sequences (calculated by SignalP 4.1) are in underlined and bold text. Highlighted in yellow is the minimal functional form of *B. subtilis* and *B. amyloliquefaciens* TapA. The imperfect repeats identified by TRUST analysis, and investigated previously, are shown in blue text (Romero et al., 2014). In blue and underlined text is the predicted β-strand region analysed in the current work. Highlighted in red is the predicted β-strand-rich region which is thought to form a stable core region resistant to proteolysis and spontaneous degradation. (Bpara = *B. paralicheniformis*, Bpum = *B. pumilis*, Bsub = *B. subtilis* and Bamy = *B. amyloliquefaciens*).

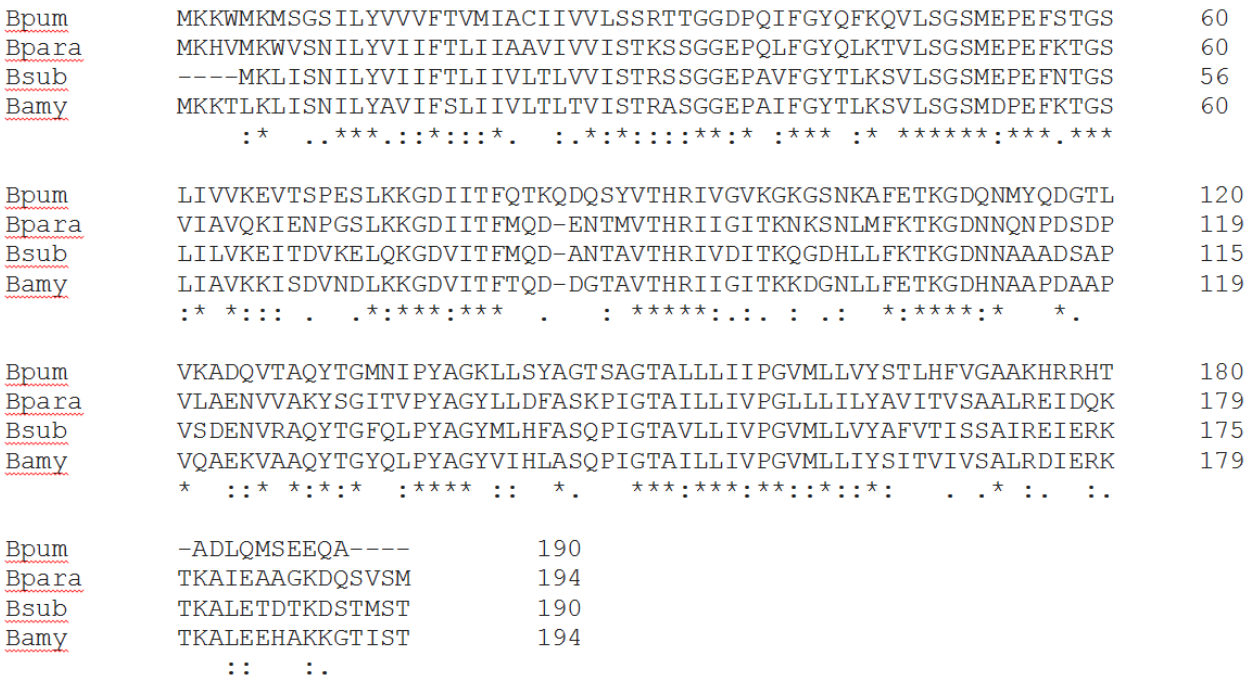


Figure S4: Multiple sequence alignment of SipW. (Bpara = *B. paralicheniformis*, Bpum = *B. pumilis*, Bsub = *B. subtilis* and Bamy = *B. amyloliquefaciens*).

Figure S5: Multiple sequence alignment of TasA. (Bcer = *B. cereus*, Bpara = *B. paralicheniformis*, Bpum = *B. pumilis*, Bsub = *B. subtilis* and Bamy = *B. amyloliquefaciens*).

Figure S6: *In silico* transcriptional terminator prediction. The online server ARNold was used to predict Rho–independent transcription terminators. The numbering refers to the numbering of the *tapA* operon where 1 = the first base of the *tapA* ORF. The terminators at 1158 and 1195 are found within the *sipW* ORF and the terminator at 2223 is found after the *tasA* sequence.

References

Romero, D., Vlamakis, H., Losick, R., and Kolter, R. (2014). Functional Analysis of the Accessory Protein TapA in *Bacillus subtilis* Amyloid Fiber Assembly. Journal of Bacteriology *196*, 1505-1513.

SCIENTIFIC REPORTS

OPEN

Natural variations in the biofilm-associated protein BslA from the genus *Bacillus*

Received: 18 November 2016
Accepted: 19 June 2017
Published online: 27 July 2017

Ryan J. Morris[1], Marieke Schor[1], Rachel M. C. Gillespie[2], Ana Sofia Ferreira[2], Lucia Baldauf[1], Chris Earl[1], Adam Ostrowski[2], Laura Hobley[3], Keith M. Bromley[1], Tetyana Sukhodub[2], Sofia Arnaouteli[2], Nicola R. Stanley-Wall[2] & Cait E. MacPhee[1]

BslA is a protein secreted by *Bacillus subtilis* which forms a hydrophobic film that coats the biofilm surface and renders it water-repellent. We have characterised three orthologues of BslA from *Bacillus amyloliquefaciens*, *Bacillus licheniformis* and *Bacillus pumilus* as well as a paralogue from *B. subtilis* called YweA. We find that the three orthologous proteins can substitute for BslA in *B. subtilis* and confer a degree of protection, whereas YweA cannot. The degree to which the proteins functionally substitute for native BslA correlates with their *in vitro* biophysical properties. Our results demonstrate the use of naturally-evolved variants to provide a framework for teasing out the molecular basis of interfacial self-assembly.

Many bacterial species live in multi-cellular communities known as biofilms[1,2]. The Gram-positive bacterium *Bacillus subtilis* forms biofilms with a characteristic wrinkled morphology and a surface that is highly hydrophobic. A critical component in generating the hydrophobicity of the biofilm is the amphiphilic protein BslA[3]. The structure of BslA contains a hydrophobic, surface-exposed 'cap' that sits atop a hydrophilic immunoglobulin-like domain, and is reminiscent of the fungal hydrophobins[3–6]. We recently characterised the mechanism by which BslA remains both stable and monomeric in aqueous environments and how its adsorbs to an interface[7]. We showed that the hydrophobic cap of BslA, in contrast to the hydrophobins, is structurally plastic; it undergoes an environmentally-responsive conformational change upon partitioning from an aqueous to a hydrophobic environment. Once at an interface, BslA self-assembles into an ordered rectangular 2D lattice that macroscopically forms an elastic protein film[3,7,8]. Like the hydrophobins, BslA has been recognized as a 'natural' surface active agent that could find a range of applications in the food and personal care industries e.g. as surface modifier, coating agent, emulsifier, and/or foam stabiliser[9–13].

Although the gross structural changes that lead to the surface-stabilising function of BslA are understood, we wished to tease out the mechanism by which BslA interacts with a hydrophilic-hydrophobic interface, i.e. what triggers the structural change between the conformational forms? Moreover, what interactions are critical for the formation of the regular 2D lattice? To address these questions, we have studied several BslA orthologues that are present in the genome of *Bacillus* species closely related to *B. subtilis*. In addition, we have investigated a paralogue of BslA called YweA. Using genetic modification and biophysical techniques, we characterize the behaviour of these variants both *in vivo* and *in vitro*. We find that key serine residues in the cap regions of the variants act as 'switches' that drive conformational re-arrangement at an interface. We propose a classification system for BslA and its variants based upon the behaviour of the protein films under compression. This categorization allows us to tease out the interface-protein and protein-protein interactions and suggest amino acids important for the formation of stable, highly organised 2D films.

Results

Protein Alignment & Choice of BslA Variants. As previously reported by Kobayashi & Iwano, a number of *Bacillus* species possess well-conserved genes that encode for homologous proteins that fall into two groups: BslA-like and YweA-like[8]. Figure 1 shows a phylogenetic tree generated via maximum likelihood analysis of the

[1]School of Physics & Astronomy, University of Edinburgh, Edinburgh, UK. [2]Division of Molecular Microbiology, School of Life Sciences, University of Dundee, Dundee, UK. [3]School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, UK. Correspondence and requests for materials should be addressed to C.E.M. (email: cait.macphee@ed.ac.uk)

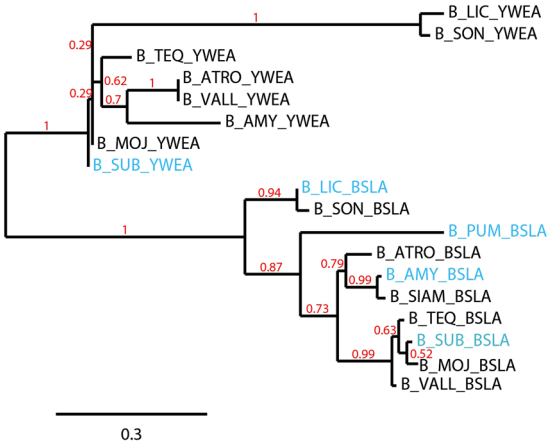


Figure 1. Maximum likelihood phylogeny of BslA and YweA protein sequences from *Bacillus* species. Abbreviations used are as follows: B_teq, Bacillus tequilensis; B_sub, Bacillus subtilis; B_pum, Bacillus pumilis; B_amy, Bacillus amyloliquefaciens; B_lic, Bacillus licheniformis; B_son, Bacillus sonorensis; B_moj, Bacillus mojavensis; B_vall, Bacillus vallismortis; B_atro, Bacillus atrophaeus; B_siam, Bacillus siamensis. The tree was rooted using the midpoint method. The bootstrap value shown in red indicates the support for each branch (from 0 to 1). YweA sequences for *Bacillus siamensis* and *Bacillus pumilus* could not be found using BlastP. Sequences highlighted in blue were used in this study.

two paralogues YweA and BslA from various *Bacillus* species. The paralogue YweA shares 67% sequence similarity with BslA and is distinguished from BslA by its lack of an N-terminal domain after the signal sequence in addition to the absence of 10 amino acids at the C-terminus that contain a conserved 'CxC' motif (Fig. 2A). In this work we have chosen to study the BslA orthologues produced by three species, *Bacillus amyloliquefaciens* (Ba_BslA), *Bacillus licheniformis* (Bl_BslA), and *Bacillus pumilus* (Bp_BslA), as well as the BslA paralogue YweA from *B. subtilis*. These particular *Bacillus* species were chosen as representatives of each distinct branch of the maximum likelihood tree (Fig. 1). The amino acid sequences of these proteins share a high degree of similiarity to *B. subtilis* BslA (Bs_BslA), while still possessing variations that may influence their behaviour both *in vivo* and *in vitro* (Fig. 2A; see Fig. S1 for the amino acid alignments for all species in the maximum likelihood tree).

In vivo characterization of BslA variants. First, we tested whether the orthologous genes, and therefore proteins, could substitute for Bs_BslA *in vivo* in *B. subtilis*. This was achieved using heterologous expression of complemented orthologous *bslA* genes (encoding for Ba_BslA, Bl_BslA, and Bp_BslA) in a mutant of *B. subtilis* possessing a *bslA* gene deletion (*bslA*$^-$). Resultant biofilm formation was assessed by four criteria: complex colony morphology, pellicle formation, colony surface hydrophobicity, and sporulation ability. Wild-type *B. subtilis* strain NCIB3610 produces characteristic biofilm and pellicle morphologies with a pronounced surface hydrophobicity, reflected in the high contact angle a water droplet makes with respect to the biofilm surface (Fig. 2D,E left column). The biofilms heterologously expressing Ba_BslA showed the greatest similarity to wild-type or Bs_BslA complemented biofilms, with a characteristic wrinkled morphology and pronounced surface hydrophobicity. The *B. subtilis* biofilms substituted with Bl_BslA were also highly hydrophobic and wrinkled, however, the wrinkles in the central 'disc' of the colony were less pronounced and the extent of wrinkling in the pellicle was reduced. For Bp_BslA, the colony surface showed marked hydrophobicity, however colony growth appeared diminished relative to NCIB3610 and the *bslA*$^-$ mutant genetically complemented with Bs_BslA. Moreover, like the *bslA*$^-$ mutant genetically complemented with Bl_BslA, the wrinkles in the central 'disc' of the colony were less pronounced. Finally, attempts to genetically complement the *bslA*$^-$ mutant with the *yweA* coding region did not reinstate either the characteristic wrinkled morphology or surface hydrophobicity of biofilms, consistent with YweA having only a minor role in the biofilm architecture (Fig. 3 and ref. 8).

The ability to sporulate is a sign of a mature biofilm matrix environment[14]. In *B. subtilis*, the *sspB* gene encodes for a small acid-soluble spore protein and expression is used as an indication of biofilm maturity. We utilized flow cytometry in conjunction with a P_{sspB}:yfp transcriptional reporter fusion to assess the degree of sporulation within each of the heterologously complemented biofilms. All heterologously expressed BslA orthologues were found to confer high levels of sporulation relative to a negative control, albeit not to the degree exhibited by the wild-type strain (Table 1). Taken together, these data indicate that the orthologous proteins Ba_BslA, Bl_BslA, and Bp_BslA can at least partially fulfil the role of Bs_BslA within the *B. subtilis* biofilm, largely reinstating surface hydrophobicity and gross morphological characteristics. It is notable, however, that Bl_BslA and, especially, Bp_BslA do exhibit some differences in biofilm morphology. YweA, in contrast, does not recover wild-type biofilm morphology or hydrophobicity (Figs 2 and 3). Nonetheless, YweA does appear to contribute to the surface hydrophobicity, exacerbating the impact of the *bslA* deletion on biofilm formation (Fig. 3).

Biophysical assessment of protein behaviour _in vitro_. Having studied the BslA variants *in vivo*, we next characterised their behaviour in isolation using *in vitro* biophysical techniques. BslA variant proteins were expressed using standard techniques (see Experimental Procedures). The resulting mature regions of the purified

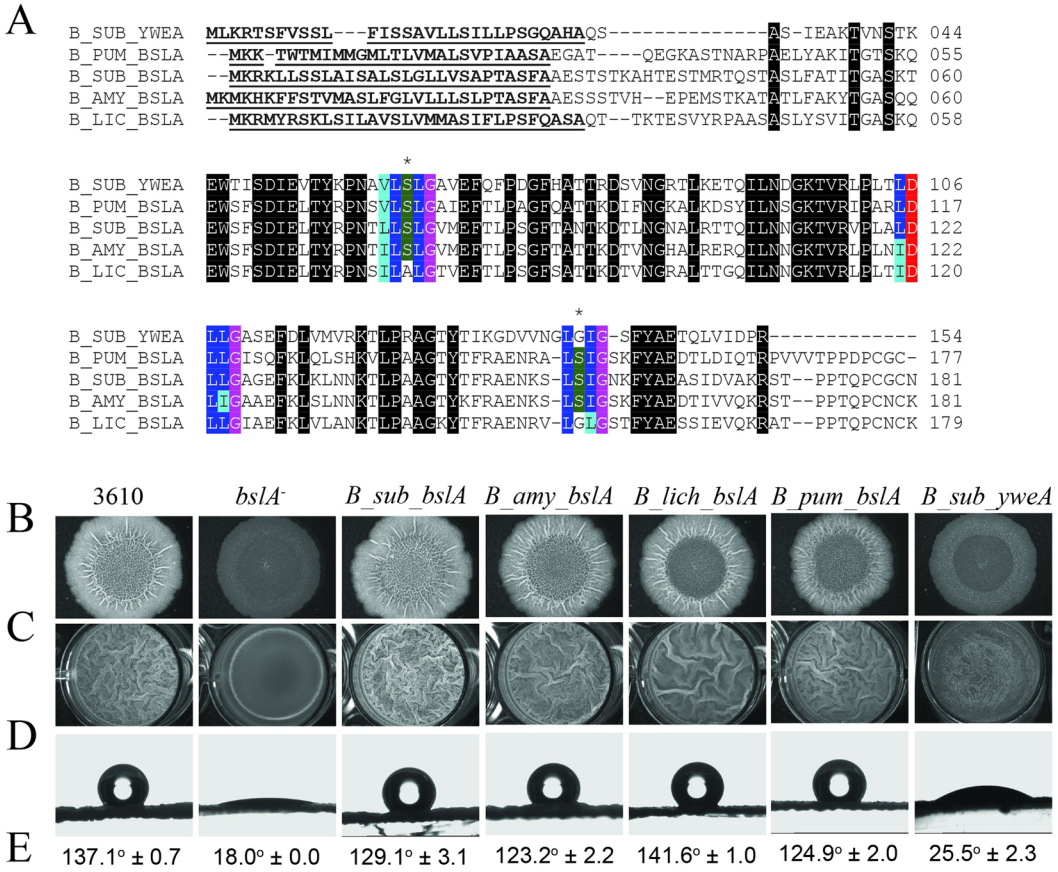


Figure 2. Assessing biofilm formation of heterologously expressed *bslA* variants in *B. subtilis*. (**A**) Amino acid alignment of BslA variants found in other *Bacillus* species. Abbreviations used are as follows: B_sub, *Bacillus subtilis*; B_pum, *Bacillus pumilis*; B_amy, *Bacillus amyloliquefaciens*; B_lic, *Bacillus licheniformis*. Underlined and bolded amino acids signify the signal sequence, black represents 100% sequence identity. Blue amino acids represent the hydrophobic cap regions, where dark blue are conserved amino acids and light blue are conservative substitutions. The purple highlighted amino acids are glycine which are conserved across all species and always follow the amino acids comprising the caps. Aspartic acid is shown in red and serine residues within caps 1 and 2 are shown in green. Note that YweA is differentiated from Bs_BslA and the BslA orthologues by the fact it lacks both the N-terminal region following the signalling sequence and the C-terminal domain. The * symbols indicate the cap regions containing serine residues. Biofilms and pellicles were formed from the wild type strain (NCIB3610), a strain possessing a *B. subtilis bslA* gene deletion mutant (*bslA*$^{-}$), strains that heterologously express the *bslA* gene of each species studied (*B. subtilis*, B_sub_bslA; *B. amyloliquefaciens*, B_sub_amy; *B. licheniformis*, B_lich_bslA; *B. pumilus*, B_amy_bslA) complemented into the *bslA*$^{-}$ mutant strain, and a strain expressing the *yweA* gene complemented into the *bslA*$^{-}$ mutant. Biofilm phenotypes were characterized by assessing (**B**) complex colony morphology; (**C**) pellicle formation; (**D**,**E**) colony surface hydrophobicity (see Table S3 for details on strains).

proteins were first analyzed using size exclusion chromatography and SDS-PAGE (Fig. 4). As demonstrated in previous work[3], purified Bs_BslA is primarily composed of monomers and dimers, with a small sub-population of higher order oligomeric species (Fig. 4A,F). The orthologous BslA proteins showed similar monomer/dimer fractions as Bs-BslA, but higher order oligomers were absent. YweA, in contrast, is almost entirely monomeric (Fig. 4E).

Kinetics of interfacial self-assembly of BslA variants. After protein purification, we assessed the interfacial activity of the BslA variants using pendant drop tensiometry. The dynamics of interfacial protein adsorption can often be characterized by three kinetic regimes: Regime I is a 'lag time' where there is no apparent change in the interfacial tension, Regime II is when a sufficient proportion of protein adsorbs to the interface to produce a decrease in interfacial tension, and Regime III is reached when the interfacial tension plateaus to a roughly constant final value[15]. Importantly, it should be stressed that once an elastic film has formed there is no longer an interfacial tension between the vapour and liquid phases and the concept of an 'interfacial tension' no longer applies. A good indication of when a film has formed is an increase in the error in the fit of the Young-Laplace equation to the drop shape[16]. Bromley *et al.* showed that the only sensible data that can be extracted from interfacial tension measurements of Bs_BslA was the time at which the interfacial tension transitions between Regime I and Regime II (the 'Regime I time'), since at any time after this transition an elastic film is present on the droplet

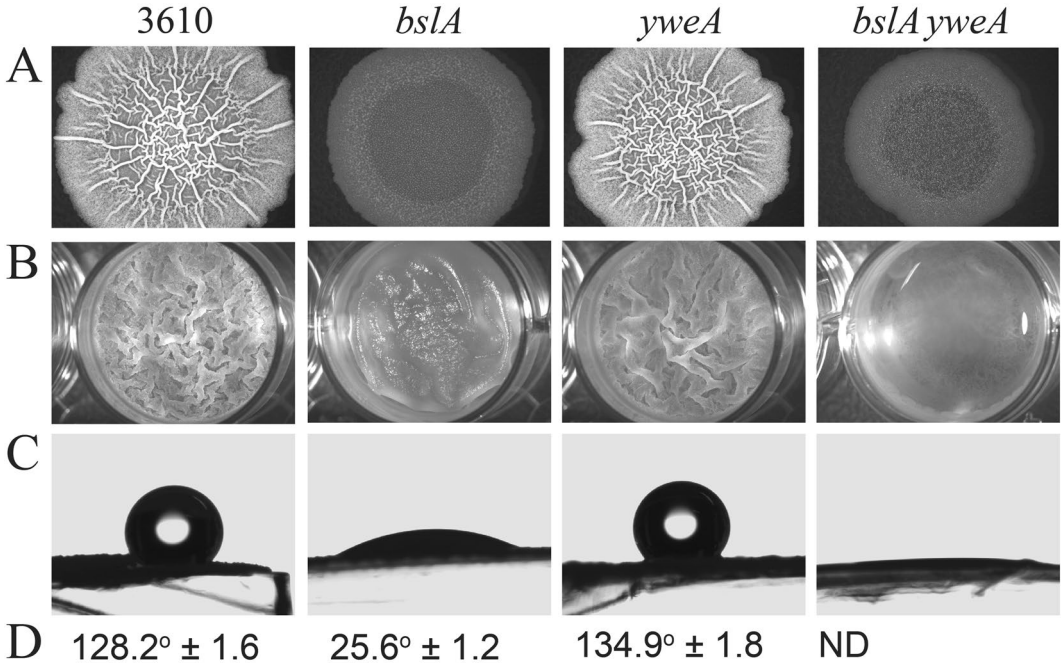


Figure 3. Characterization of biofilm phenotypes of *yweA* mutants. Biofilm phenotypes were characterized by assessing (**A**) complex colony morphology; (**B**) pellicle formation; (**C**,**D**) colony surface hydrophobicity for *B. subtilis* wild-type strain 3610, a *bslA* mutant, a *yweA* mutant, and a double mutant *bslA yweA* (see Table S3). Row C shows a representative sessile drop resting atop a biofilm formed from one of the strains used above. Row D are the average contact angles; ND means that a contact angle could not be determined from the images.

Strain	% Sporulation positive[a]	
	0 μM IPTG	25 μM IPTG
3610 P*sspB-yfp*	26.6	24.8 ± 1.63
bslA P*sspB-yfp*	5.9 ± 0.6	4.9 ± 0.8
bslA P_{IPTG}-$bslA_{B.sub}$ P*sspB-yfp*	3.2 ± 0.2	14.1 ± 2.3
bslA P_{IPTG}-$bslA_{B.lic}$ P*sspB-yfp*	5.6 ± 0.3	15.6 ± 1.2
bslA P_{IPTG}-$bslA_{B.amy}$ P*sspB-yfp*	5.6 ± 0.4	13.0 ± 3.7
bslA P_{IPTG}-$bslA_{B.pum}$ P*sspB-yfp*	5.2 ± 0.2	15.6 ± 1.2

Table 1. Sporulation frequency assessed by flow cytometry. [a]The percentage sporulation was calculated using fluorescence form the P*sspB-ypf* fusion as a proxy after 48 hours incubated at 30 °C. NCIB3610 was used as the non-fluorescent control. For the full genotypes of the strains see Table S3. For each condition at least two independent samples were analysed. Expression of the variant gene was achieved using 25 μM IPTG. The value presented is the mean and the error is the standard deviation.

surface[7]. Regime I times were determined in one of two ways: (1) the transition time between regimes I and II when the fit error was still low (<0.4 μm) or (2) when the fit error increased to a threshold value (>0.75 μm). Criterion (1) or (2) was chosen by whichever occurred first. We found that the Regime I times of the orthologues Ba_BslA and Bp_BslA were within error of Bs_BslA (Fig. 5A). In contrast, the Regime I time of Bl_BslA was nearly twice as long as those of the other orthologues while the Regime I time for YweA was faster than that of Bs_BslA by ~25%.

Relaxation dynamics of elastic films. The formation of an elastic film at a hydrophobic/hydrophilic interface is demonstrated by the observation of wrinkles following compression of the droplet (achieved by withdrawing a small volume of fluid following equilibration). One way of characterizing these films is by monitoring the formation and subsequent disappearance of these wrinkles over time, reflecting relaxation of the interfacial protein film (Fig. 5B). We have assumed that wrinkle relaxation is due to the loss of protein into the subphase, although rearrangement of the protein at the interface cannot be ruled out. For these experiments an oil/water interface was used as the small difference in density between the two phases results in rounder droplets, allowing for better imaging of the wrinkles that form. As found previously[7], wrinkles formed under compression in a Bs_BslA film do not relax over the timescale of the experiment. In contrast, Ba_BslA and Bl_BslA exhibit very

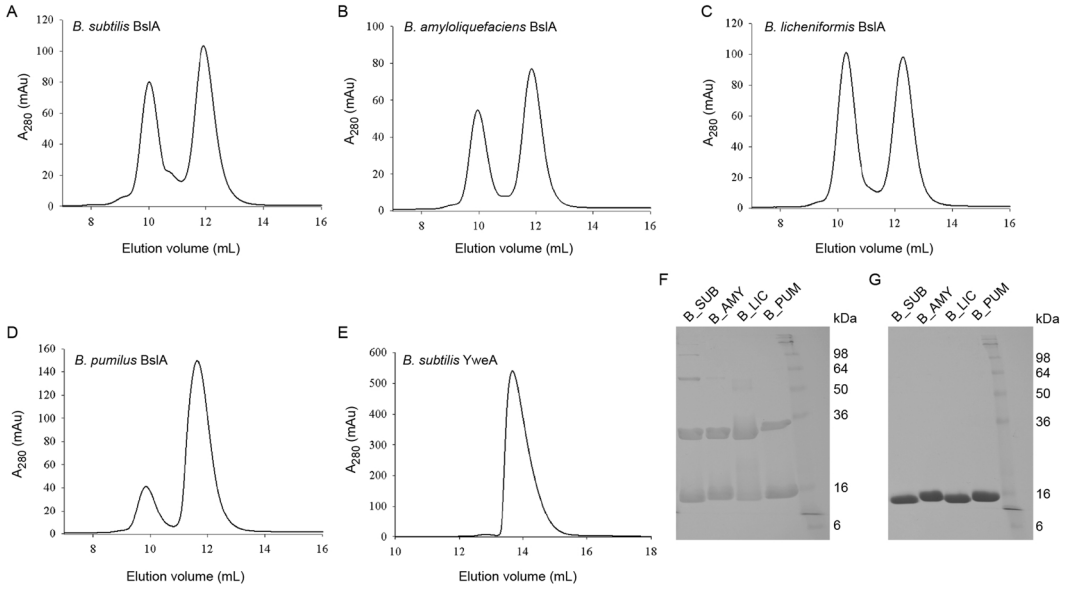


Figure 4. Size exclusion chromatography and SDS-PAGE analysis of Monomer-Dimer Formation. The *bslA* variant genes were cloned, overexpressed as a GST fusion and then purified. (**A**) Using size exclusion chromatography Bs_BslA was found to contain two primary peaks corresponding to monomer and dimer species, although there are higher-order additional peaks present. (**B**) Ba_BslA, (**C**) Bl_BslA, and (**D**) Bp_BslA all eluted in primarily two peaks. (**E**) YweA, in contrast, eluted in effectively one monomer peak. (**F**) SDS-PAGE analysis of recombinant protein without β-mercaptoethanol and (**G**) with β-mercaptoethanol.

slow relaxation, and Bp_BslA relaxes within a minute. Strikingly, relaxation of YweA is extremely rapid, occurring immediately upon compression.

Film organisation of BslA variants. Bs_BslA forms highly ordered 2D rectangular lattices at an air/water interface[7]. We investigated the structure of the films formed by BslA variants using transmission electron microscopy (TEM). Protein lattices were observed for all variants, but exhibited varying degrees of order. The films formed by Ba_BslA showed domains of ordered protein, but these were not widespread (Fig. 5C). As shown by the Fast Fourier Transform (FFT), weak peaks can be observed, indicating some level of ordering (Fig. 5C). The Bl_BslA films exhibited larger domains of ordered protein, and the FFT shows clear peaks (Fig. 5D). In contrast, Bp_BslA showed very weak ordering with only very small patches of organized structure, similar to those previously observed for the BslA L77K mutant[3] (Fig. 5E). Thus amongst BslA orthologues, there appears to be some correlation between wrinkle relaxation and microscopic ordering: Bs_BslA films are highly ordered and wrinkles formed under compression do not relax, Ba_BslA and Bl_BslA show some degree of ordering and display slow wrinkle relaxation dynamics, whilst Bp_BslA films are not well ordered and wrinkles formed upon compression relax rapidly.

The paralogue YweA, however, does not conform to the trend observed for BslA orthologues. Indeed, YweA formed clearly defined and well ordered films similar to those formed by Bs_BslA, despite its rapid relaxation (Fig. 5B,F), calling into question the relationship between the microscopic organization of the protein within a film and film robustness.

CD spectroscopy of solution and interfacial states. Finally, circular dichroism (CD) spectroscopy was used to study the conformation of the BslA variants in aqueous solution and at an oil/water interface (Fig. 5G). Qualitatively, the spectra of the different BslA variants in aqueous solution are similar, with only Bp_BslA showing the loss of a feature between 210 and 218 nm (Fig. 5G). When bound to refractive indexed matched emulsions (RIMEs; Fig. 5H) all of the proteins underwent a structural transition consistent with an increase in β-sheet structure, reflecting rearrangement of the hydrophobic cap upon insertion into the oil phase.

Effects of Bl_BslA cap mutations on the properties of *in vitro* protein films and *in vivo* biofilm morphology. Pendant drop tensiometry showed that it takes Bl_BslA approximately twice as long as the other orthologues to become adsorbed at an interface (Fig. 5A). We have previously interpreted this slower-than-diffusion adsorption time in relation to the energy barrier between solution and interfacial conformations of BslA[7], implying that there is a larger energy barrier that must be overcome for Bl_BslA adsorption. Comparison of the amino acid sequence of Bs_BslA with that of Bl_BslA reveals two amino acid differences in the hydrophobic cap region (Fig. 2A). Specifically, the first difference resides in 'Cap 1' (residues L77, S78, and L79 of Bs_BslA) in which serine-78 in Bs_BslA is replaced by an alanine (A76 in Bl_BslA). The second difference is in 'Cap 3' (residues L153, S154, and I155 of Bs_BslA) in which S154 is replaced by a glycine (G152 in Bl_BslA). In contrast, these two serine residues are conserved in Ba_BslA and Bp_BslA, which have similar adsorption

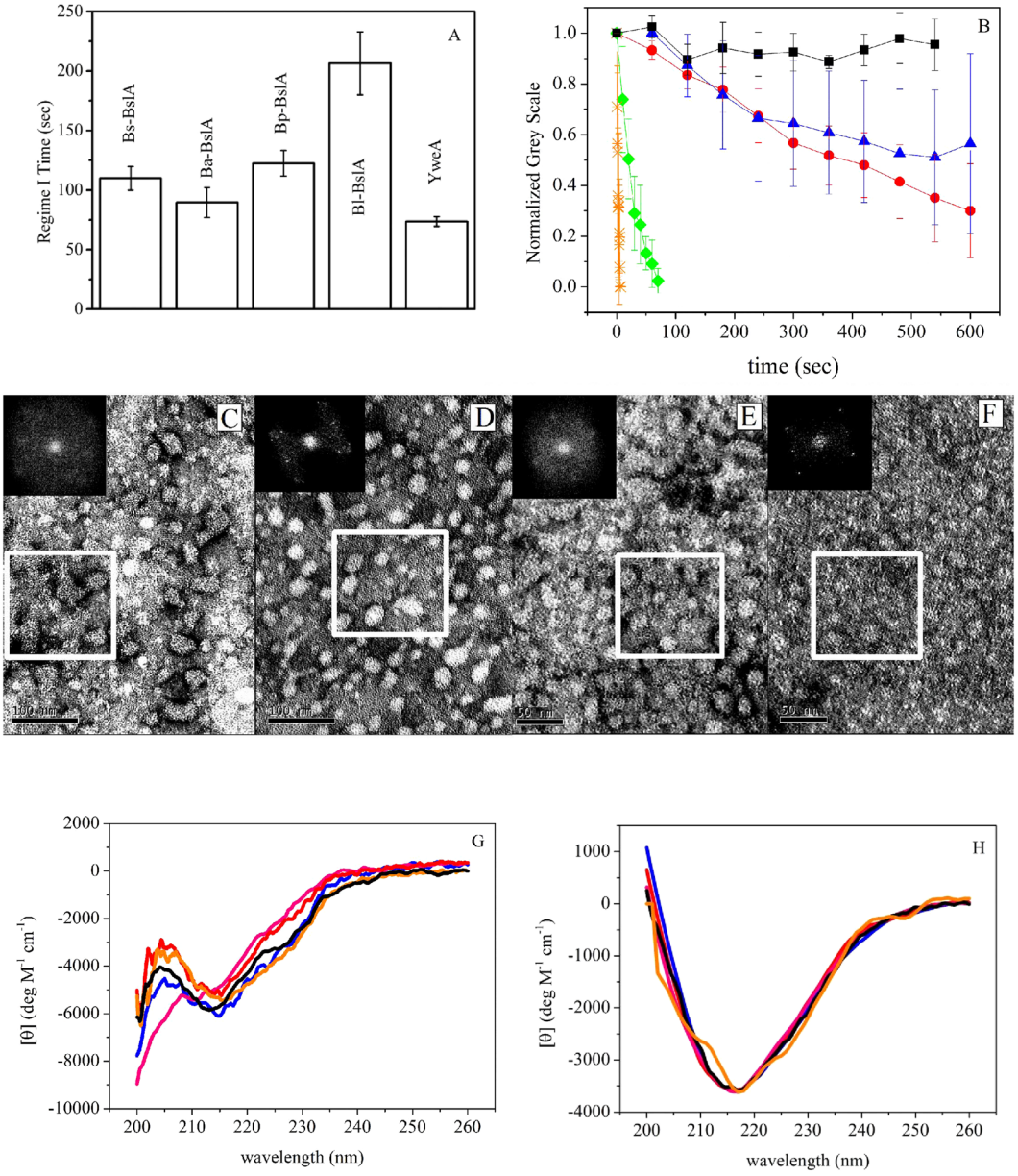


Figure 5. Biophysical characterization of BslA orthologues. (**A**) Regime I times were measured using pendant drop tensiometry. Plotted is the mean of 4 separate repeat experiments and the error bars correspond to the standard deviation. (**B**) Wrinkles were formed in the protein films by compressing the pendant droplet via removal of fluid. The relaxation of these wrinkles are plotted as a function of time: Bs_BslA(black squares); Ba_BslA (blue triangles), Bl_BslA (red circles); Bp_BslA (green diamonds); YweA (orange stars). Error bars represent standard deviation in relaxation times for at least N = 10 separate wrinkles across the drop. (**C**) Inset in each image is the FFT contained within the white box. Ba_BslA visually showed ordered domains, but was not widespread.The FFT displays weak peaks indicating some ordering. (**D**) Bl_BslA has perceptibly larger domains of order. The FFT contains obvious peaks indicating ordering of the protein. (**E**) TEM of Bp_BslA films showed very weak ordering with patchy organization, which is reflected in the isotropic FFT. (**F**) YweA formed the ordered films that were the most similar to Bs_BslA, as can be seen from the FFT. (**G**) Solution state circular dichroism spectra of BslA variants and (**H**) circular dichroism spectra of RIMEs: Bs_BslA (black); Ba_BslA (blue); Bl_BslA (red); Bp_BslA (green); YweA (orange). Note the RIME CD spectra were normalized assuming all protein is adsorbed to emulsion droplets.

kinetics to the Bs_BslA. To investigate the role of these residues in the kinetics of adsorption and the properties of the resultant film, we produced two single amino acid mutants of Bl_BslA (A76S and G152S) and the double mutant A76S/G152S in which the cap region is identical to that of Bs_BslA. If these cap regions, and their re-orientation at an interface, determine the kinetics of adsorption, one would predict that the Regime I time of Bl_BslA(A76S/G152S) would coincide with those of Bs_BslA and the other orthologues. This is in fact what is observed as shown in Fig. 6B. In contrast, both single mutants result in Regime I times that are shorter than those

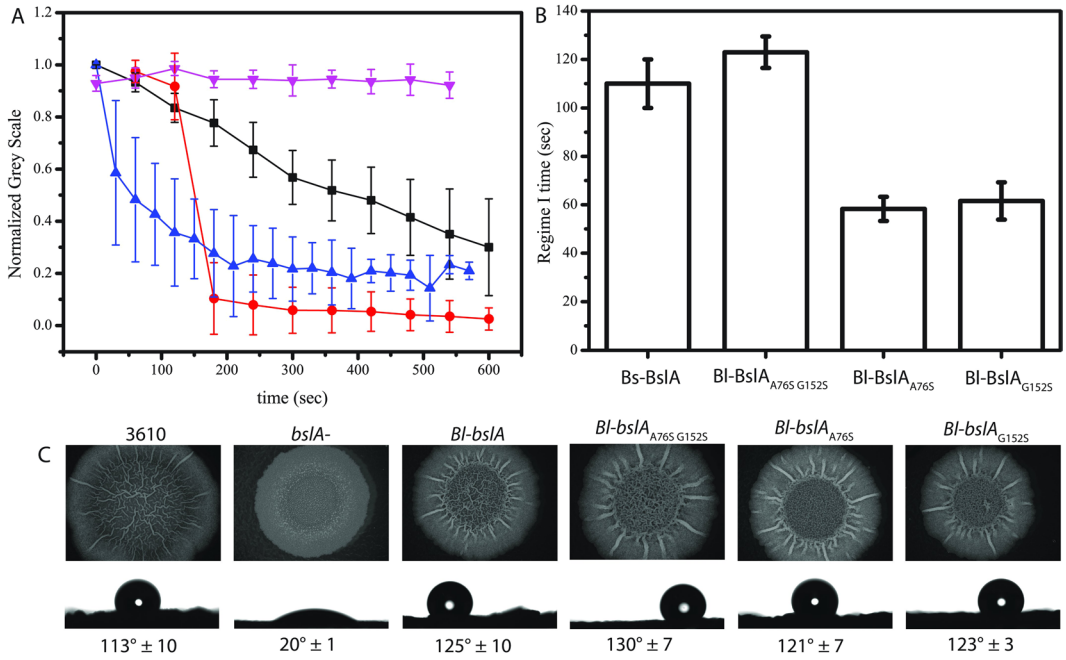


Figure 6. *In vitro* and *in vivo* effects of the Bl_BslA cap mutations. (**A**) Wrinkle relaxation measurements for Bl_BslA (black squares), Bl_BslA(A76S) (red circles), Bl_BslA(G152S) (blue triangles), Bl_BslA(A76S G152S) (upside down pink triangles). Error bars represent standard deviation in relaxation times for at least N = 10 separate wrinkles across the drop. (**B**) Regime I times of Bs_BslA (repeated from Fig. 3 for clarity) and the Bl_BslA single and double mutants. (**C**) Heterologous expression of Bl_BslA variants shows no deleterious effects on biofilm morphology or hydrophobicity relative to both wild-type Bs_BslA (as determined using the parental NCIB3610 strain (3610)) and parental Bl_BslA constructs.

of Bs_BslA and Bl_BslA(A76S/G152S) but very similar to that of YweA. Interestingly, the 'Cap 3' region of YweA possesses the same S → G substitution as Bl_BslA (Fig. 2A). The fact that the single mutants are more 'YweA-like' in terms of adsorption indicates that residues 76 and 152 of Bl_BslA help shape the energetic landscape of cap re-organization at an interface. To further explore this hypothesis, we also studied the relaxation of the films formed by these mutant proteins (Fig. 6A). Notably, wrinkles formed under compression of a Bl_BslA(A76S/G152S) droplet did not relax over the time scale of the experiment, unlike Bl_BslA which relaxes slowly, lending further support to a role for these cap residues in determining the stability and mechanical properties of the film. Consistent with this idea, the single cap mutants more closely recapitulate YweA-like behaviour, with films formed by these proteins exhibiting more rapid relaxation than both Bl_BslA and Bl_BslA(A76S/G152S) films. Indeed, films formed by Bl_BslA(A76S), the amino acid composition of which most closely resembles that of YweA, demonstrates the fastest relaxation rate of the cap mutants investigated. Intriguingly, the relaxation of wrinkles formed by this mutant appears to plateau to a non-zero value over the course of the experiment, indicating that wrinkles persist. We conclude from these results that the serine/glycine residue found in Cap 3 of BslA variants plays a prominent role in both the energetics of adsorption and the robustness of the resultant film.

Finally, we tested how the Bl_BslA cap mutations affect the structure and properties of the *B. subtilis* biofilm. When the genes encoding the Bl_BslA cap variants are heterologously expressed, the resultant biofilm morphology and hydrophobicity are indistinguishable from both wild-type Bs_BslA (as determined using the parental NCIB3610 strain) and parental Bl_BslA constructs (Fig. 6C). Thus, while the *in vitro* films produced from these mutants show distinctive physical characteristics, we do not observe any differences in *in vivo* phenotype.

Discussion

While the structural metamorphosis of the hydrophobic cap has been identified as the mechanism that allows BslA to be stable in both aqueous and hydrophobic environments, many questions remain regarding how BslA self-assembles into 2D ordered lattices and what the key amino acids are that contribute to network formation at interfaces. We have previously demonstrated that single amino acid mutations can have marked consequences for both the kinetics and energetics of adsorption along with the spatial ordering of the protein at the interface[7,17]. Instead of using a brute-force site-directed mutagenesis approach to further illuminate the biophysical properties of BslA at an interface, we have investigated the BslA variants found in three other *Bacillus* species, along with the *B. subtilis* BslA paralogue YweA. The aim of this strategy is to lead to more efficient targeting of important amino acids for future systematic mutagenesis studies of Bs_BslA. Such knowledge could be exploited to tailor both the kinetics of interfacial adsorption as well as the mechanical properties of the protein film for the application of BslA in multiphase formulations.

We have confirmed that YweA plays no obvious role in overall biofilm formation and morphology *in vivo*, as was demonstrated in previous studies[8]. However, the phenotype of the double deletion mutant *bslA yweA*

indicates that YweA is indeed present within the biofilm and makes a small but noticeable contribution to surface hydrophobicity[8] (Fig. 3). The major difference between YweA and the other BslA variants is the lack of both the N- and C-terminal domains. Since YweA cannot recover the characteristics of the wild-type biofilm phenotype, it is possible that the lack of one or both of these domains is functionally important for biofilm formation, in either mediating interactions between other BslA proteins, or interactions with other protein or polysaccharide components of the biofilm matrix.

Importantly, the biophysical characterization of YweA interfacial activity and self-assembly provides new insights into the relationship between the microscopic organization of the protein at the interface and resulting macroscopic properties of the surface layer. YweA interfacial films showed extremely rapid relaxation under applied compression (Fig. 5B), yet YweA films imaged by TEM showed the greatest order of all the variants (Fig. 5F). In previous studies, the cap mutant Bs_BslA(L77K) also formed films that relaxed within very short time scales. However, this variant demonstrated significant 2D disorder[3,7], and we suggested that the single mutation in the cap disrupted the ability of Bs_BslA(L77K) to form a space-spanning network by altering the orientation of the protein at an interface[17]. The behaviour of YweA runs counter to this hypothesis, indicating that film order does not necessarily correlate with film robustness. It can be concluded, then, that YweA has the ability to form interfacial protein-protein lateral interactions but that these are weak, particularly compared to those formed by Bs_BslA.

As Bl_BslA differed markedly in its adsorption kinetics relative to the other BslA variants investigated here, this orthologue formed the focus of further analyses. The amino acids of Bl_BslA differ from Bs_BslA and the other variants by two non-conservative substitutions in Cap 1 (S → A) and Cap 3 (S → G). Examination of the interfacial conformation of BslA reveals the two serines to be immediately adjacent to one another, both orientated towards the interior of protein (Fig. S1), and are likely to form a sidechain-sidechain hydrogen bond. In contrast, when the protein is in aqueous solution and the cap region is disordered, these serines are oriented outward towards the aqueous environment. The switch from an outward to an inward orientation would facilitate the conformational switch of the cap region upon contact with a hydrophobic interface, with hydrogen bonding between the residues further stabilising the interfacial conformation.

The Bl_BslA cap is more hydrophobic than Bs_BslA and the other orthologues - the polar serines are replaced by nonpolar amino acids (A76 and G152). Counter-intuitively, the higher hydrophobicity of the cap region overall does not result in faster adsorption (Fig. 5A). The resolution of this apparent paradox resides in the fact that the cap is a flexible, plastic structure. In an aqueous environment, the more hydrophobic residues of the Bl_BslA cap will most likely be oriented inwards, away from the aqueous phase, whereas in Bs_BslA the equivalent hydrophilic side chains are oriented outwards. We propose that the driving force for cap rearrangement in Bs_BslA is reorientation of these serines away from the hydrophobic interface and that this mechanism is lost in Bl_BslA, resulting in slower adsorption. Thus, we hypothesize that it is the hydrophilic residues within the cap that act as a 'sensor' for hydrophobic interfaces. Interaction with an apolar interface induces a reorientation of the hydrophilic amino acid(s), which in turn triggers the larger scale reorganization of the cap such that the hydrophobic residues become surface-exposed and subsequently restructure to form a three-stranded β-sheet.

Interestingly, the single cap mutants of Bl_BslA show shorter adsorption times, very similar to the Regime I time of YweA (Figs 5A and 6B). YweA has a similar amino acid cap content as the Bl_BslA(A76S) mutant, with a serine in Cap 1 and a glycine in Cap 3. For Bl_BslA(G152S), there is a hydrophobic residue in Cap 1 and a hydrophilic residue in Cap 3. We hypothesize that this combination of a hydrophilic residue in one cap strand, and a hydrophobic residue in the other, may cause a tension within the aqueous conformation of the cap as a whole or, alternatively, facilitate a rapid interconversion between states. This tension or interconversion then primes the cap to more easily undergo the conformational change when it encounters an interface, effectively lowering the energy barrier to re-structuring. Moreover, when in the interfacial conformation, the remaining serine will be unfavourably partitioned into the hydrophobic protein core, but without a second serine present with which to form a hydrogen bond. This may explain why the proteins are displaced from the interface under compression.

Taking the result of this work together, we propose a categorization of BslA and its variants based upon their wrinkle relaxation behaviour:

- **Surfactant-like**: The surfactant-like proteins comprise YweA, Bp_BslA and the Bs_BslA mutants L77K and L79K[3]. Whilst YweA differs from the other three BslA variants in its ability to form ordered lattices at an interface, wrinkles formed under compression of all four proteins relax within very short time scales. This implies that the interfacial lateral protein-protein interactions either cannot form correctly due to incorrect orientation at the interface (in the case of L77K and L79K[17]) or are weak due to substitution of critical amino acids.
- **Transiently wrinkling film formers**: The proteins that belong to this category are Ba_BslA, Bl_BslA, Bl_BslA(A76S), and Bl_BslA(G152S). In these cases, wrinkles persist for longer time scales than those formed by the surfactant-like proteins, but do eventually relax over time. Such behaviour suggests relatively weak lateral interactions between proteins at the interface, but stronger than those exhibited by the surfactant-like proteins.
- **Rigid film formers**: Bs_BslA and Bl_BslA(A76S + G152S) are members of this group. The stresses imposed via our methods are insufficient to dislodge protein from the interface, and moreover the entire droplet shape distorts under compression. This implies that there exist not only strong protein-interface interactions but well established protein-protein interactions within the film.

Categorizing the BslA variants in this manner allows us to identify links between the *in vivo* morphology of *Bacillus* biofilms (Figs 3 and 4) and the interactions the proteins experience at an interface. Thus the extent to which each BslA variant rescues the wild type *B. subtilis* biofilm may be ranked and mapped to the classification system outlined above (Table 2). Bs_BslA produces rigid, elastic films *in vitro* and creates the highly wrinkled

Rank	Protein	Film category
1	Bs_BslA	Rigid film
2	Ba_BslA	Transient wrinkling film
3	Bl_BslA	Transient wrinkling film
4	Bp_BslA	Surfactant-like
5	YweA	Surfactant-like

Table 2. Ranking of biofilm phenotype vs *in vitro* film classification.

and hydrophobic biofilm phenotype associated with *B. subtilis* pellicles and colonies. Ba_BslA and Bl_BslA share similar film characteristics *in vitro* and in both cases the biofilm morphology is similar to the *B. subtilis* wild-type phenotype with some small, but discernible differences. Bp_BslA falls within the surfactant-like class and is the least effective of the orthologues at recovering the wild-type *B. subtilis* biofilm morphology. Finally, the paralogue YweA is the most surfactant-like of the proteins investigated and clearly cannot recover any features of the biofilm phenotype. Relating the biofilm morphology to the classes of film formed *in vitro* shows a clear correlation between film robustness and biofilm structure.

This classification provides useful insights into which residues are important for protein-protein interactions within the film. A large proportion of outward-facing, non-cap residues are conserved across all four variants. However, amino acids that are conserved between Bl_BslA and Bs_BslA but not between these two proteins and Ba_BslA and Bp_BslA are likely to affect the strength of the interaction between the proteins in the film. This suggests roles for S49, T105, Q107, N111 and E128 in the formation of the optimal lateral interactions that give rise to stable films (Fig. 7). Specifically, comparing the amino acid sequences of each orthologue reveals that Bp_BslA possesses non-conservative substitutions in four out of these five positions, Ba_BslA has one non-conservative substitution, and Bl_BslA is identical to Bs_BslA. This correlates well with our observations of interfacial protein lattice structure (Fig. 5C–F). The subtle difference in lattice structure between Bl_BslA and Bs_BslA, despite the conservation of these putative lateral interaction sites, we hypothesize originates from the key serine residues (A76S, G152S) identified in this work. These residues may influence how the protein is tilted at an interface, which may in turn influence the extent and strength of lateral interactions between adjacent proteins. We have previously modeled how changes in the cap region influence average protein tilt[17].

Further *in vivo* investigation will be required to determine whether the biophysical differences measured as part of this study have evolved specifically to generate a fitness benefit in the distinct environmental habitat or niche occupied by each species of bacteria.

Experimental Procedures

Bacterial strains and growth conditions. *Escherichia coli* and *Bacillus subtilis* strains used and constructed in this study are detailed in Table S1. All strains were routinely grown in Lysogeny broth (LB; 10 g NaCl, 5 g yeast extract and 10 g tryptone per litre) or on LB solidified with 1.5% (w/v) select agar (Invitrogen) at 37 °C. *B. subtilis* biofilms were grown in MSgg medium (5 mM potassium phosphate and 100 mM MOPS at pH 7.0 supplemented with 2 mM $MgCl_2$, 700 μM $CaCl_2$, 50 μM $MnCl_2$, 50 μM $FeCl_3$, 1 μM $ZnCl_2$, 2 μM thiamine, 0.5% (v/v) glycerol, and 0.5% (v/v) glutamate). When required, antibiotics were used at the following concentrations: 100 $\mu g\,ml^{-1}$ ampicillin, 100 $\mu g\,ml^{-1}$ spectinomycin, 25 $\mu g\,ml^{-1}$ kanamycin, 1 $\mu g\,ml^{-1}$ erythromycin and 25 $\mu g\,ml^{-1}$ lincomycin. Ectopic gene expression was induced with 25 μM isopropyl β-D-1 thiogalactopyranoside (IPTG). *E. coli* strain MC1061 [*F'lacIQ lacZM15 Tn10* (*tet*)] was used for the routine construction and maintenance of plasmids. *B. subtilis* 168 derivatives were generated by transformation of competent cells with plasmids using standard protocols[18]. SPP1 phage transductions, for the introduction of DNA into *B. subtilis* strain NCIB3610, were performed as previously described[19].

Plasmid construction. The primers and plasmids used in this study are presented in Tables S2 and S3 respectively. Single amino acid substitutions were generated by PCR site-directed mutagenesis using KOD Hot Start DNA Polymerase (Novagen) and the appropriate primer pairs from Table S2, with reaction conditions calculated according to the Stratagene manual for QuikChange Site-Directed Mutagenesis. The resulting mutant plasmids were digested with DpnI and transformed into competent *E. coli* MC1061 cells. All mutations were verified by DNA sequencing. For expression of BslA orthologues in *B. subtilis* under control of the IPTG-inducible Phyper-spank promoter, the *bslA* gene was amplified from genomic DNA isolated from *Bacillus licheniformis* DSM13; *Bacillus amyloliquefaciens* FZB42; and *Bacillus pumilus* SAFR-032 using the primer pairs NSW812/NSW813; NSW829/NSW830; and NSW819/NSW820 respectively. Resulting PCR products were digested with HindIII and SphI, and ligated into the *B. subtilis* shuttle vector, pDR111 (also digested HindIII/SphI), to enable subsequent gene integration into the non-essential *amyE* locus of the *B. subtilis* chromosome (final pDR111-derived plasmids are listed in Table S3). Expression of *yweA* was achieved in a similar manner. The *yweA* gene was amplified from *B. subtilis* NCIB3610 genomic DNA using primers NSW810 and NSW811, digested HindIII/SphI and ligated into HindIII/SphI-digested pDR111.

Plasmid pNW1420, for over-expression of $YweA_{31-155}$ in *E. coli*, was generated as follows: $yweA_{31-155}$ was amplified from *B. subtilis* NCIB3610 genomic DNA using primers NSW1853 and NSW1854 and ligated into the expression vector pET15bTEV using the NdeI and XhoI restriction sites. The resulting plasmid encodes the N-terminally His_6-tagged fusion protein, His_6-$YweA_{31-155}$, which contains the recognition site of Tobacco Etch Virus (TEV) protease between His_6 and $YweA_{31-155}$ coding regions, allowing removal of the His_6 tag by digestion

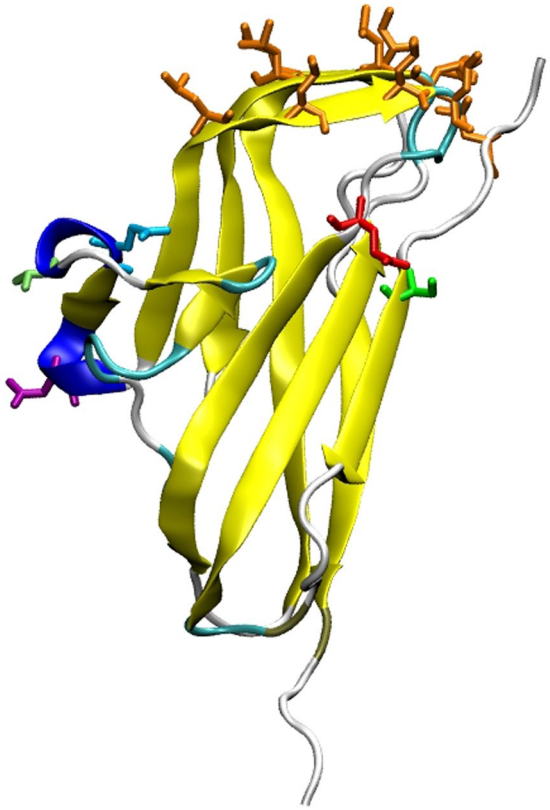


Figure 7. Possible amino acids that mediate interfacial protein-protein interactions. 3D structure of BslA highlighting the hydrophobic residues involved in interfacial adsorption (orange) and the residues identified as putatively involved in protein-protein interaction S49, T105, Q107, N111 and E128 (green = S, light green = T, cyan = Q, purple = N, red = E).

with TEV. Plasmids for production of BslA orthologues from *B. amyloliquefaciens*, *B. licheniformis*, and *B. pumilus* encoded truncated proteins corresponding to *B. subtilis* $BslA_{42–181}$ (determined by sequence alignment; see Fig. 2A), on which previous analyses were based[3,8], and were generated as follows. For over-expression of *B. amyloliquefaciens* $BslA_{42–181}$ (referred to herein as Ba_BslA), the appropriate DNA fragment was amplified from *B. amyloliquefaciens* FZB42 genomic DNA using primers NSW2011 and NSW2012, and ligated into the GST expression vector pGEX-6-1P via BamHI and XhoI sites. The resulting plasmid, pNW1422, encodes the N-terminally tagged fusion protein GST-Ba_BslA, from which the GST tag can be removed by TEV protease digestion Plasmid pNW1423 for over-expression of *B. licheniformis* $BslA_{40–179}$ (Bl_BslA) and pNW1424 for over-expression of *B. pumilus* $BslA_{37–177}$ (Bp_BslA), were generated in an identical manner using primer pairs NSW2013/NSW2014 and *B. licheniformis* DSM13 genomic DNA, and NSW2015/NSW2016 and *B. pumilus* SAFR-032 genomic DNA respectively. All resulting constructs were sequence-verified.

Biofilm analysis. Biofilm analysis was performed as previously described, with minor modifications[19]. For complex colony formation, 10 μL of the appropriate *Bacillus* culture, grown to mid-exponential phase in LB, was spotted onto MSgg medium solidified with 1.5% (w/v) agar and incubated at 30 °C for 48 hours. For pellicle analysis, the LB starter culture was diluted 1:100 (v/v) in MSgg medium and incubated at 25 °C for 72 hours. Both colony and pellicle biofilms were imaged using a Leica MZ16 stereoscope (Leica Microsystems).

Colony hydrophobicity contact angle measurements. Biofilms were grown on MSgg agar plates (as described above). A section through the mature biofilm was extracted using a scalpel and placed on a microscope slide. A 5 μl droplet of sterile double-distilled water was placed on the colony using the ThetaLite Optical Tensiometer with OneAttension software. The drop was allowed to equilibrate for 5 minutes prior to imaging and contact angle measurement.

Sporulation assays and FACS analysis. Strains harbouring the promoter *PspoB* fused with *yfp* coding sequence were used to assess sporulation levels in mature colonies by flow cytometry[20]. The experiments used biofilms grown for 2 days at either 30 °C on MSgg agar without IPTG or supplemented with 25 μM IPTG. After incubation, biofilms were collected, washed and fixed with 4% paraformaldehyde (PFA), as previously described[14]. Fluorescence was analysed at the single-cell level using a BD FACSCalibur (BD Biosciences)[21].

Protein Purification. BslA proteins were purified as previously described[7], with minor changes as follows. *E. coli* BL21 (DE3) pLysS cells were transformed with the appropriate plasmids for over-expression of GST-BslA fusion proteins (described above; Table S3). For protein production, the transformed cells were grown in autoinduction medium[22] supplemented with ampicillin (100 μg mL^{-1}) at 37 °C and 200 rpm until an OD_{600} of 0.9 was reached, at which point protein expression was induced by reducing the temperature to 18 °C for further incubation overnight. Cells were collected by centrifugation at 4000 × *g* for 45 min and stored at −80 °C until further use. For purification, bacterial pellets were thawed and resuspended in purification buffer (50 mM HEPES pH 7.5, 300 mM NaCl (for YweA), or 50 mM HEPES pH 7.5, 250 mM NaCl (for all other purifications)) supplemented with Complete EDTA-free Proteinase Inhibitors (Roche) before lysis using an Emulsiflex cell disruptor (Avestin). Unlysed cells and cell debris were removed by centrifugation at 27000 × *g* for 20 min, and the cleared lysates incubated with Glutathione Sepharose 4B resin (GE Healthcare; for BslA) or Ni-nitrilotriacetic acid (Ni-NTA) agarose (Qiagen; for YweA) at a ratio of 750 μL resin per 1 L bacterial culture with gentle rotation for 4 h at 4 °C. To isolate the bound fusion proteins, the mixture of lysate plus beads was passed through a single-use 25-mL gravity flow column (Bio-Rad), and the collected beads washed twice with 20 mL of the appropriate purification buffer. Untagged BslA variants were generated by resuspending the washed beads in 25 mL purification buffer supplemented with 1 mM DTT and 0.5 mg TEV protease prior to gentle rotation overnight at 4 °C. To generate untagged YweA from the His_6-YweA fusion protein, the overnight mixture was additionally supplemented with 250 mM imidazole. Following overnight cleavage, mixtures were again passed through gravity flow columns and the flow-through collected. To the flow-through, 750 μL fresh Glutathione Sepharose plus 250 μL Ni-NTA agarose (for BslA variants), or 100 μL Ni-NTA agarose (for YweA), was added and the solution incubated with gentle rotation overnight at 4 °C to remove the TEV protease and any unbound GST or His_6. The mixtures were passed through the gravity flow columns for a final time, and purified proteins collected in the flow-through. The purified proteins were then concentrated (with simultaneous exchanges into 25 mM phosphate buffer pH 7.0 where appropriate) and further purified by size exclusion chromatography (SEC) as previously described[7].

SDS-PAGE analysis. SDS-PAGE analysis of BslA variants was performed using 30 μg purified samples of BslA diluted 4:1 (v/v) in 4X loading buffer (6.2 g SDS, 40 mL 0.5 M Tris pH 6.8, 6.4 mL 0.1 M EDTA, 32 mL glycerol, 1 mg Bromophenol blue) either with or without β-mercaptoethanol at 4% (v/v). Samples to which β-mercaptoethanol was added were boiled at 100 °C for 5 min prior to analysis, whilst samples without β-mercaptoethanol were not subjected to boiling. Proteins were run on a standard 14% polyacrylamide denaturing gel at 200 V for 60 min and visualised by staining with Coomassie Blue.

Mass spectrometry analysis. Purified proteins were identified by excision of protein bands from SDS-PAGE gels and LC-MS-MS following tryptic digest. Protein size estimation was performed using MS-TOF and peptides identified from the MASCOT database. All procedures were performed by FingerPrints Proteomics service, School of Life Sciences, Dundee.

Bioinformatics. BslA orthologues were identified using BLASTP[23,24] using the non-reduntant protein sequence database for each of the organisms of interest using the protein sequence of BslA from *B. subtilis* as the query. BslA orthologues were distinguished from YweA based on the presence of a C-terminal region not present in YweA[8] (see Fig. 2A and Table S4). BslA and YweA protein sequences were aligned using Clustal Omega with the default settings[25]. The signal sequences were predicted using the SignalP 4.1 server[26] and the aligned sequences were manually coloured to highlight the 'cap' regions and identity. A maximum likelihood tree was calculated from the Clustal Omega alignment using phylogeny.fr's standard workflow[27], where Gblocks[28] was used to eliminate divergent and poorly aligned columns, and a tree computed with PhyML[29] using a WAG substitution model[30] and 100 bootstrap replicates. The outputted tree was visualised using TreeDyn[31].

Pendant drop tensiometry & and surface wrinkle relaxation quantification. Pendant drop experiments were performed on a Krüss EasyDrop tensiometer. Protein samples were diluted in buffer and immediately placed in a syringe with a needle diameter of 1.83 mm. Images of the pendant drop are captured by a CCD camera and Krüss software fits the Young-Laplace equation to the drop shape to determine the interfacial tension. For measuring the dynamic interfacial tension, samples were prepared by diluting each protein in phosphate buffer to a concentration of 0.03 mg ml^{-1}. Droplets were expelled in air and the interfacial tension was determined by fitting the droplet shape to the Young-Laplace equation. For the film relaxation experiments, a protein concentration of 0.2 mg ml^{-1} was used. A droplet of the protein in aqueous solution was expelled into glyceryl trioctanoate oil and allowed to equilibrate at room temperature for 30 minutes. Images were acquired at 2 fps using a digital camera. Wrinkle relaxation experiments were performed following compression of the droplet, which was achieved by retracting 5 μL of the protein solution. Compression induced the formation of wrinkles in the surface layer. The wrinkles were monitored over a 10 minute period. To analyze the relaxation of the wrinkles, a line profile was drawn across the wrinkles. The line profile was plotted using the greyscale values (from 0 to 255) of each pixel along this line using the ImageJ. At least 10 wrinkles were monitored over the time course of the experiment. To plot the relaxation rate, the greyscale value of the pixels was normalized and background corrected.

Transmission electron microscopy (TEM). Protein samples were deposited onto carbon-coated copper grids (Cu-grid) (TAAB Laboratories Equipment Ltd) and imaged using a Philips/FEI CM120 BioTwin transmission electron microscope. A 5 μL droplet of protein (0.025 mg mL^{-1}) was pipetted onto a grid and allowed to equilibrate for 5 mins before being wicked with filter paper from the side. A 5 μL droplet of 2% uranyl acetate was

then applied to the grid and similarly left for 5 min before being wicked from the side. The Fast Fourier Transform (FFT) of the images was performed using ImageJ software.

Circular dichroism analysis. Circular dichroism spectropolarimetry (CD) was performed using a Jasco J-810 spectropolarimeter. Samples were analysed at a concentration of 0.03 mg mL^1 in a 1 cm quartz cuvette. Measurements were performed with a scan rate of 50 nm sec^{-1}, a data pitch of 0.1 nm and a digital integration time of 1 sec. Twenty accumulations were measured and averaged to produce the final curve. Refractive index matched emulsions (RIMEs) were made by first preparing a 20% (v/v) decane emulsion with 0.2 mg ml^{-1} of protein. The emulsion was mixed for 1 min using a rotor stator (IKA Ultra-Turrax T10) at 30,000 RPM. The emulsion was washed three times in order to remove any residual protein not adsorbed to an oil/water interface. Washes were performed by centrifuging at 1000 rpm for 20 sec, a portion of subphase was removed and replaced with buffer, then gently re-dispersed. Finally, subphase was removed and replaced with glycerol such that the final wt% of glycerol was 59% (w/v). The emulsion was then gently remixed on a rollerbank and then allowed to cream. The cream was placed in a 1 mm pathlength quartz cuvette for spectrum measurement. To prevent creaming during the experiment, the cuvette was briefly inverted between measurements to re-disperse the droplets.

References

1. Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R. & Lappin-Scott, H. M. Microbial biofilms. *Annual Reviews in Microbiology.* **49**(1), 711–745 (1995).
2. Hall-Stoodley, L., Costerton, J. W. & Stoodley, P. Bacterial biofilms: from the natural environment to infectious diseases. *Nature Reviews Microbiology.* **2**(2), 95–108 (2004).
3. Hobley, L. *et al*. BslA is a self-assembling bacterial hydrophobin that coats the *Bacillus subtilis* biofilm. *Proceedings of the National Academy of Sciences.* **110**(33), 13600–13605 (2013).
4. Linder, M. B., Szilvay, G. R., Nakari-Setälä, T. & Penttilä, M. E. Hydrophobins: the protein-amphiphiles of filamentous fungi. *FEMS Microbiology Reviews.* **29**(5), 877–896 (2005).
5. Hakanpää, J. *et al*. Two crystal structures of *Trichoderma reesei* hydrophobin HFBI–the structure of a protein amphiphile with and without detergent interaction. *Protein Science.* **15**(9), 2129–2140 (2006).
6. Torkkeli, M., Serimaa, R., Ikkala, O. & Linder, M. Aggregation and self-assembly of hydrophobins from *Trichoderma reesei*: low-resolution structural models. *Biophysical Journal.* **83**(4), 2240–2247 (2002).
7. Bromley, K. M. *et al*. Interfacial self-assembly of a bacterial hydrophobin. *Proceedings of the National Academy of Sciences.* **112**(17), 5419–5424 (2015).
8. Kobayashi, K. & Iwano, M. BslA(YuaB) forms a hydrophobic layer on the surface of *Bacillus subtilis* biofilms. *Molecular Microbiology.* **85**(1), 51–66 (2012).
9. Valo, H. K. *et al*. Multifunctional hydrophobin: toward functional coatings for drug nanoparticles. *ACS Nano.* **4**(3), 1750–1758 (2010).
10. Hektor, H. J. & Scholtmeijer, K. Hydrophobins: proteins with potential. *Current Opinion in Biotechnology.* **16**(4), 434–439 (2005).
11. Tchuenbou-Magaia, F. L., Norton, I. T. & Cox, P. W. Hydrophobins stabilised air-filled emulsions for the food industry. *Food Hydrocolloids.* **23**(7), 1877–1885 (2009).
12. Green, A. J., Littlejohn, K. A., Hooley, P. & Cox, P. W. Formation and stability of food foams and aerated emulsions: Hydrophobins as novel functional ingredients. *Current Opinion in Colloid & Interface Science.* **18**(4), 292–301 (2013).
13. Reger, M., Sekine, T., Okamoto, T., Watanabe, K. & Hoffmann, H. Pickering emulsions stabilized by novel clay–hydrophobin synergism. *Soft Matter.* **7**(22), 11021–11030 (2011).
14. Vlamakis, H., Aguilar, C., Losick, R. & Kolter, R. Control of cell fate by the formation of an architecturally complex bacterial community. *Genes & Development.* **22**(7), 945–953 (2008).
15. Beverung, C. J., Radke, C. J. & Blanch, H. W. Protein adsorption at the oil/water interface: characterization of adsorption kinetics by dynamic interfacial tension measurements. *Biophysical Chemistry.* **81**(1), 59–80 (1999).
16. Alexandrov, N. A. *et al*. Interfacial layers from the protein HBFII hydrophobin: Dynamic surface tension, dilatational elasticity and relaxation times. *Journal of Colloid & Interface Science.* **376**(1), 296–306 (2012).
17. Brandani, G. B. *et al*. The bacterial hydrophobin BslA is a switchable ellipsoidal janus nanocolloid. *Langmuir.* **31**(42), 11558–11563 (2015).
18. Harwood, C. R. & Cutting, S. M. Molecular biological methods for *Bacillus* (Wiley, 1990).
19. Verhamme, D. T., Kiley, T. B. & Stanley-Wall, N. R. DegU co-ordinates multicellular behaviour exhibited by *Bacillus subtilis*. *Molecular Microbiology.* **65**(2), 554–568 (2007).
20. Marlow, V. L. *et al*. Phosphorylated DegU manipulates cell fate differentiation in the *Bacillus subtilis* biofilm. *Journal of Bacteriology.* **196**(1), 16–27 (2014).
21. Murray, E. J., Strauch, M. A. & Stanley-Wall, N. R. σ^X is involved in controlling *Bacillus subtilis* biofilm architecture through the AbrB homologue Abh. *Journal of Bacteriology.* **191**(22), 6822–6832 (2009).
22. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression & Purification.* **41**(1), 207–234 (2005).
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology.* **215**(3), 403–410 (1990).
24. Altschul, S. F. *et al*. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Research.* **25**(17), 3389–3402 (1997).
25. Sievers, F. *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology.* **7**(1), 539 (2011).
26. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth.* **8**, 785–786 (2011).
27. Dereeper, A. *et al*. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* **36**, W465–W469 (2008).
28. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution* **17**, 540–552 (2000).
29. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* **52**, 696–704 (2003).
30. Whelan, S. & Goldman, N. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and Evolution* **18**, 691–699 (2001).
31. Chevenet, F., Brun, C., Bañuls, A., Jacq, B. & Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**, 439–439 (2006).

Acknowledgements

We would like to acknowledge the Flow Cytometry and Cell Sorting Facility and the FingerPrints Proteomics Facility at the University of Dundee. We would also like to thank Dr. Jim B. Procter for help with the bioinformatics. This work has been supported by funding from the Engineering and Physical Sciences Research Council [EP/J007404/1] and the Biotechnology and Biological Sciences Research Council [BB/L006979/1; BB/I019464/1; BB/L006804/1].

Author Contributions

R.J.M. designed research, performed experiments & data analysis, wrote & edited manuscript; M.S. performed experiments & data analysis, edited manuscript; R.M.C.G. performed experiments & data analysis, edited manuscript; A.S.F. provided experimental materials, performed experiments & data analysis; L.B. performed experiments & data analysis; C.E. performed experiments & data analysis; A.O. performed experiments; L.H. performed experiments; K.M.B. performed experiments & data analysis; T.S. provided experimental materials; S.A. performed experiments; N.S.W. designed research, wrote & edited manuscript; C.E.M. designed research, wrote & edited manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-06786-9

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.